Quasi-optimal and adaptive sparse grids with control variates for PDEs with random diffusion coefficient

F. Nobile[†], L. Tamellini[♯], R. Tempone[♭], F. Tesei[†]

[†] CSQI - MATHICSE, EPFL, Switzerland
[♯] Dipartimento di Matematica, Università di Pavia, Italy
[♭] SRI UQ Center, KAUST, Saudi Arabia

UQAW 2016
KAUST, January 5, 2016

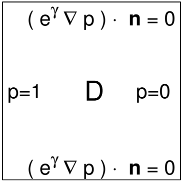
Outline

The uncertain Darcy problem

Find a pressure $p : \overline{D} \times \Gamma \to \mathbb{R}$, such that

$$\begin{cases} -\nabla \cdot (e^{\gamma_{\nu}} \nabla p) = f & \text{in } D, \\ +B.C. \text{ (see plot on the right).} \end{cases}$$

u = outward flux from the right-hand boundary.

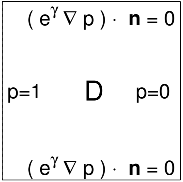


The uncertain Darcy problem

Find a pressure $p(\mathbf{x}, \mathbf{y}) : \overline{D} \times \Gamma \to \mathbb{R}$, such that ϱ-a.e.:

$$\begin{cases} -\nabla \cdot (e^{\gamma_{\nu}(\mathbf{x},\mathbf{y})} \nabla p(\mathbf{x}, \mathbf{y})) = f(\mathbf{x}) & \mathbf{x} \in D, \\ +B.C. \text{ (see plot on the right).} \end{cases}$$

$u(\mathbf{y})$ = outward flux is a random fun.$\to$ appr. e.g. $\mathbb{E}[u]$



γ_{ν} = rand. field with **tensor Matérn covariance** expanded by Kar.-Loève

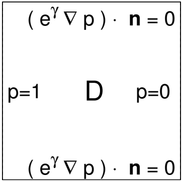
$$\gamma_{\nu}(\mathbf{x}, \mathbf{y}) = \sigma \sum_{k=1}^{\infty} y_k \gamma_{k,\nu} \phi_k(\mathbf{x}), \quad y_i \sim \mathcal{N}(0, 1) \text{ i.i.d.}$$

The uncertain Darcy problem

Find a pressure $p(\mathbf{x}, \mathbf{y}) : \overline{D} \times \Gamma \to \mathbb{R}$, such that ϱ-a.e.:

$$\begin{cases} -\nabla \cdot (e^{\gamma_{\nu}(\mathbf{x},\mathbf{y})} \nabla p(\mathbf{x}, \mathbf{y})) = f(\mathbf{x}) \quad \mathbf{x} \in D, \\ +B.C. \text{ (see plot on the right).} \end{cases}$$

$u(\mathbf{y})$ = outward flux is a random fun. $\to$ appr. e.g. $\mathbb{E}[u]$



γ_{ν} = rand. field with **tensor Matérn covariance** expanded by Kar.-Loève

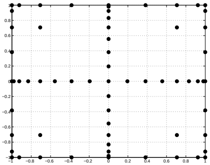
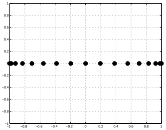
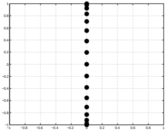
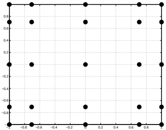
$$\gamma_{\nu}(\mathbf{x}, \mathbf{y}) = \sigma \sum_{k=1}^{\infty} y_k \gamma_{k,\nu} \phi_k(\mathbf{x}), \quad y_i \sim \mathcal{N}(0, 1) \text{ i.i.d.}$$

The smoothness of the realizations of γ (hence the decay of $\gamma_{k,\nu}$) depends on $\nu \in [0.5, \infty]$:

- $\nu = 0.5 \Rightarrow \gamma_{\nu}$ = exponential cov. fun. ($C^{0,s}, s < 1/2$ realizations)
- $\nu = \infty \Rightarrow \gamma_{\nu}$ = Gaussian cov. fun. (C^{∞} realizations)

Hierarchical representation of a sparse grid

$\mathcal{S}_{\mathcal{I}}[u](\mathbf{y}) = \sum_{\mathbf{i}\in\mathcal{I}} c_{\mathbf{i}} \mathcal{T}^{m(\mathbf{i})}[u](\mathbf{y})$ is a **sparse grid interp.**

 $=$  $+$  $+$  $\dots$

$\mathcal{S}_{\mathcal{I}}[u](\mathbf{y}) \quad = \quad \mathcal{T}^{[15\ 1]}[u](\mathbf{y}) \quad + \quad \mathcal{T}^{[1\ 15]}[u](\mathbf{y}) \quad + \quad \mathcal{T}^{[5\ 5]}[u](\mathbf{y}) \quad \dots$

$\mathcal{Q}_{\mathcal{I}}[u] = \sum_{\mathbf{i}\in\mathcal{I}} c_{\mathbf{i}} \mathbb{E}\left[\mathcal{T}^{m(\mathbf{i})}[u]\right] = \sum_{\mathbf{i}\in\mathcal{I}} c_{\mathbf{i}} \sum_{\mathbf{y}_j \in \mathcal{T}^{m(\mathbf{i})}} \omega_j u(\mathbf{y}_j)$ is a **s.g. quadrature**

Admissibility condition for $\mathcal{I}$: $\quad \forall \mathbf{i} \in \mathcal{I}, \quad \mathbf{i} - \mathbf{e}_j \in \mathcal{I}$ if $i_j > 1$.

Collocation pts: $y_i \sim \mathcal{N}(0,1) \rightarrow$ Gauss–Hermite, Genz–Keister, gen. Leja

Hierarchical representation of a sparse grid

$$\mathcal{S}_{\mathcal{I}}[u](\mathbf{y}) = \sum_{\mathbf{i}\in\mathcal{I}} c_{\mathbf{i}} \mathcal{T}^{m(\mathbf{i})}[u](\mathbf{y}) = \sum_{\mathbf{i}\in\mathcal{I}} \boldsymbol{\Delta}^{m(\mathbf{i})}[u](\mathbf{y})$$ is a **sparse grid interp.**

- $\mathcal{U}_n^{m(i_n)}[u](y_n)$ is an **interpolant** along y_n over $m(i_n)$ points
- $\mathcal{T}^{m(\mathbf{i})}[u](\mathbf{y}) = \bigotimes_{n=1}^{N} \mathcal{U}_n^{m(i_n)}[u](\mathbf{y})$ is a **tensor interpolant**
- $\mathcal{U}_n^{m(i_n)}[u] - \mathcal{U}_n^{m(i_n-1)}[u]$ is an **interpolation detail**
- $\boldsymbol{\Delta}^{m(\mathbf{i})}[u](\mathbf{y}) = \bigotimes_{n=1}^{N} (\mathcal{U}_n^{m(i_n)}[u] - \mathcal{U}_n^{m(i_n-1)}[u])$ is a **hierarchical surplus**

$$\mathcal{Q}_{\mathcal{I}}[u] = \sum_{\mathbf{i}\in\mathcal{I}} c_{\mathbf{i}} \mathbb{E}\left[\mathcal{T}^{m(\mathbf{i})}[u]\right] = \sum_{\mathbf{i}\in\mathcal{I}} c_{\mathbf{i}} \sum_{\mathbf{y}_j \in \mathcal{T}^{m(\mathbf{i})}} \omega_j u(\mathbf{y}_j)$$ is a **s.g. quadrature**

Admissibility condition for $\mathcal{I}$: $\forall\, \mathbf{i} \in \mathcal{I}, \quad \mathbf{i} - \mathbf{e}_j \in \mathcal{I}$ if $i_j > 1$.

Collocation pts: $y_i \sim \mathcal{N}(0,1) \rightarrow$ Gauss–Hermite, Genz–Keister, gen. Leja

Optimized (knapsack) sparse grids

$\mathcal{S}_{\mathcal{I}}[u](\mathbf{y}) = \sum_{\mathbf{i}\in\mathcal{I}} \mathbf{\Delta}^{m(\mathbf{i})}[u](\mathbf{y})$. **What terms $\mathbf{\Delta}^{m(\mathbf{i})}$ to include in the sum?**

1. How much will the approx. improve if I add $\mathbf{\Delta}^{m(\mathbf{i})}[u]$? $\rightarrow \Delta E(\mathbf{i})$
2. How many new problem solves if I add $\mathbf{\Delta}^{m(\mathbf{i})}[u]$? $\rightarrow \Delta W(\mathbf{i})$

$\downarrow$

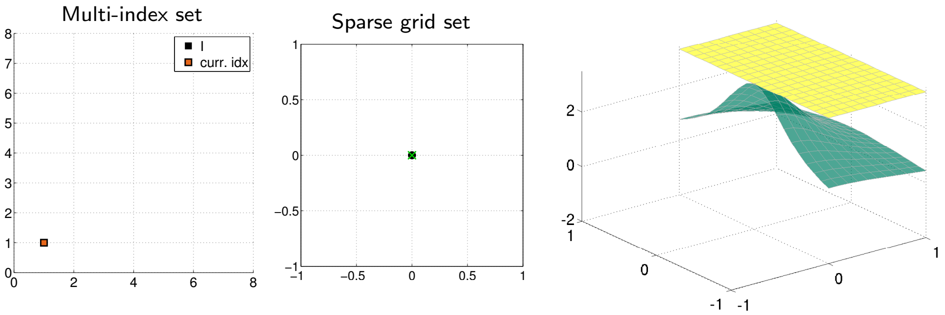
Use as $\mathcal{I}$ the set of $\mathbf{\Delta}^{m(\mathbf{i})}[u]$ with the highest ratio $\dfrac{\Delta E(\mathbf{i})}{\Delta W(\mathbf{i})}$

(i.e. a knapsack problem solved by the Dantzig algorithm more)

Note: $\Delta E(\mathbf{i})$ depends on the norm used to "measure the improvement".
We show next three different strategies, two a-posteriori and one a-priori

A-posteriori (adaptive) strategies

Gerstner & Griebel, 2003



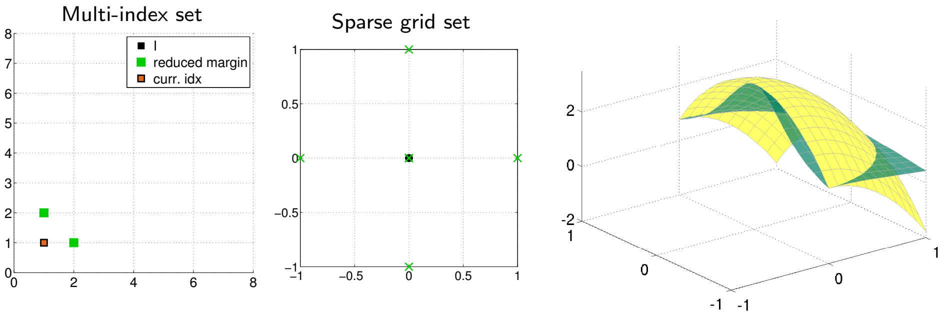
Given $\mathbf{i} = \mathbf{1}$, $\mathcal{I} = \{\mathbf{i}\}$ and $\mathcal{R} = \varnothing$ repeat:

1. Add to $\mathcal{R}$ the neighbors of $\mathbf{i}$ feasible wrt to $\mathcal{I}$
2. Compute $\mathcal{S}_{\mathcal{I}\cup\mathcal{B}}[u]$
3. find the index $\mathbf{j} \in \mathcal{R}$ with the highest profit (estimated "a-posteriori")
4. set $\mathbf{i} = \mathbf{j}$ and move it from $\mathcal{R}$ to $\mathcal{I}$

NB: slight changes needed for non-nested points and unbounded Γ, see later and poster.

A-posteriori (adaptive) strategies

Gerstner & Griebel, 2003



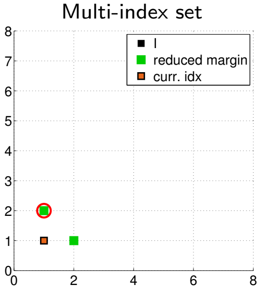
Given $\mathbf{i} = \mathbf{1}$, $\mathcal{I} = \{\mathbf{i}\}$ and $\mathcal{R} = \varnothing$ repeat:

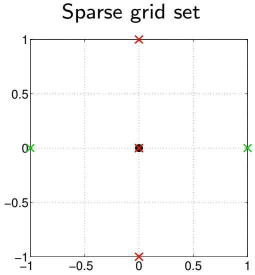
1. Add to $\mathcal{R}$ the neighbors of $\mathbf{i}$ feasible wrt to $\mathcal{I}$
2. Compute $\mathcal{S}_{\mathcal{I} \cup \mathcal{B}}[u]$
3. find the index $\mathbf{j} \in \mathcal{R}$ with the highest profit (estimated "a-posteriori")
4. set $\mathbf{i} = \mathbf{j}$ and move it from $\mathcal{R}$ to $\mathcal{I}$

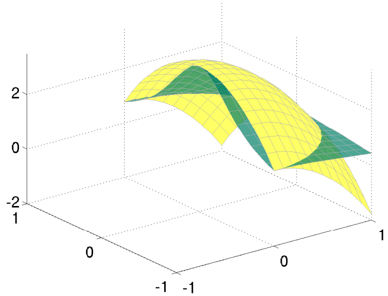
NB: slight changes needed for non-nested points and unbounded Γ, see later and poster.

A-posteriori (adaptive) strategies

Gerstner & Griebel, 2003







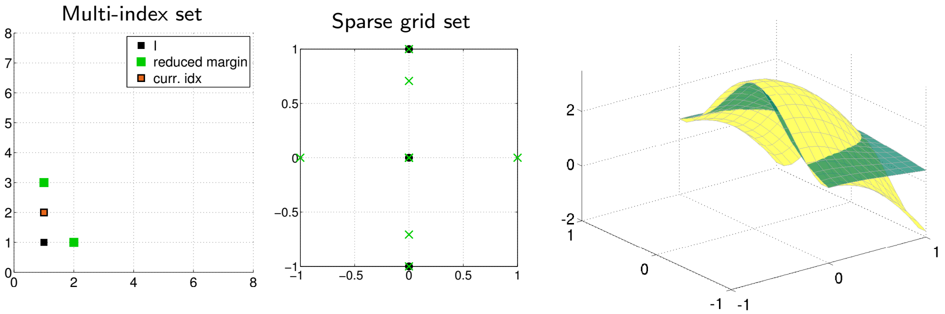
Given $\mathbf{i} = \mathbf{1}$, $\mathcal{I} = \{\mathbf{i}\}$ and $\mathcal{R} = \emptyset$ repeat:

1. Add to $\mathcal{R}$ the neighbors of $\mathbf{i}$ feasible wrt to $\mathcal{I}$
2. Compute $\mathcal{S}_{\mathcal{I}\cup\mathcal{B}}[u]$
3. find the index $\mathbf{j} \in \mathcal{R}$ with the highest profit (estimated "a-posteriori")
4. set $\mathbf{i} = \mathbf{j}$ and move it from $\mathcal{R}$ to $\mathcal{I}$

NB: slight changes needed for non-nested points and unbounded Γ, see later and poster.

A-posteriori (adaptive) strategies

Gerstner & Griebel, 2003



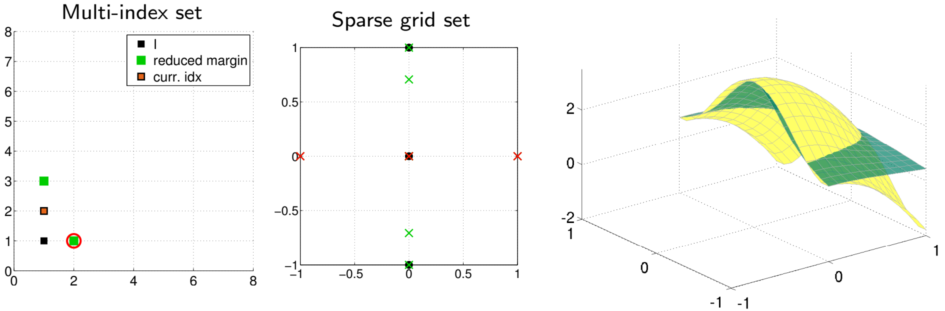
Given $\mathbf{i} = \mathbf{1}$, $\mathcal{I} = \{\mathbf{i}\}$ and $\mathcal{R} = \emptyset$ repeat:

1. Add to $\mathcal{R}$ the neighbors of $\mathbf{i}$ feasible wrt to $\mathcal{I}$
2. Compute $\mathcal{S}_{\mathcal{I} \cup \mathcal{B}}[u]$
3. find the index $\mathbf{j} \in \mathcal{R}$ with the highest profit (estimated "a-posteriori")
4. set $\mathbf{i} = \mathbf{j}$ and move it from $\mathcal{R}$ to $\mathcal{I}$

NB: slight changes needed for non-nested points and unbounded Γ, see later and poster.

A-posteriori (adaptive) strategies

Gerstner & Griebel, 2003



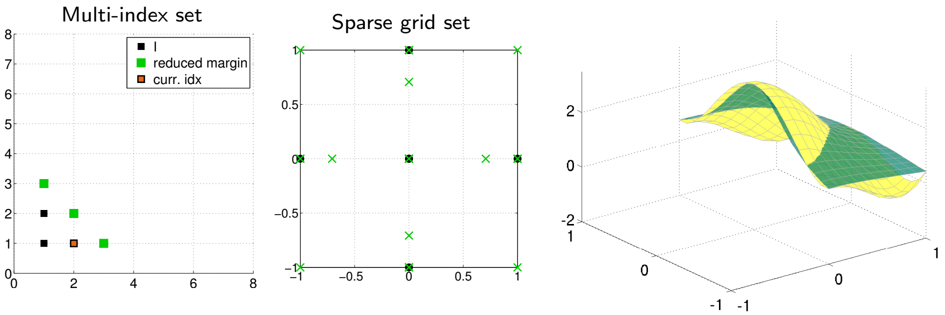
Given $\mathbf{i} = \mathbf{1}$, $\mathcal{I} = \{\mathbf{i}\}$ and $\mathcal{R} = \emptyset$ repeat:

1. Add to $\mathcal{R}$ the neighbors of $\mathbf{i}$ feasible wrt to $\mathcal{I}$
2. Compute $\mathcal{S}_{\mathcal{I} \cup \mathcal{B}}[u]$
3. find the index $\mathbf{j} \in \mathcal{R}$ with the highest profit (estimated "a-posteriori")
4. set $\mathbf{i} = \mathbf{j}$ and move it from $\mathcal{R}$ to $\mathcal{I}$

NB: slight changes needed for non-nested points and unbounded Γ, see later and poster.

A-posteriori (adaptive) strategies

Gerstner & Griebel, 2003



Given $\mathbf{i} = \mathbf{1}$, $\mathcal{I} = \{\mathbf{i}\}$ and $\mathcal{R} = \varnothing$ repeat:

1. Add to $\mathcal{R}$ the neighbors of $\mathbf{i}$ feasible wrt to $\mathcal{I}$
2. Compute $\mathcal{S}_{\mathcal{I}\cup\mathcal{B}}[u]$
3. find the index $\mathbf{j} \in \mathcal{R}$ with the highest profit (estimated "a-posteriori")
4. set $\mathbf{i} = \mathbf{j}$ and move it from $\mathcal{R}$ to $\mathcal{I}$

NB: slight changes needed for non-nested points and unbounded Γ, see later and poster.

Method 1: Quadrature-adaptive sparse grids

Here we consider $\Delta E(\mathbf{i}) \approx \mathbb{E}\left[\boldsymbol{\Delta}^{m(\mathbf{i})}[u]\right]$.

1. $$\Delta W(\mathbf{i}) = \begin{cases} \prod_{n=1}^{N} \left(m(i_n) - m(i_n - 1)\right) & \text{Genz–Keister} \\ \prod_{n=1}^{N} m(i_n) & \text{Gauss–Hermite} \end{cases}$$
   or alternatively $\Delta W(\mathbf{i}) = 1$.
2. $\Delta E(\mathbf{i}) = |\mathcal{Q}_{\mathcal{I}\cup\{\mathbf{i}\}}[u] - \mathcal{Q}_{\mathcal{I}}[u]|$ for any $\mathcal{I}$ admissible st $\mathcal{I} \cup \{\mathbf{i}\}$ is admissible.

Note: The algorithm above is not bound to use nested points! Hence we can use Gauss–Hermite points, whose $m(\cdot)$ grows much slower than Genz–Keister nodes.

Method 2: Interpolation-adaptive sparse grids

Here we consider $\Delta E(\mathbf{i}) \approx \max_{\mathbf{y}\in\Gamma}\left(\left|\boldsymbol{\Delta}^{m(\mathbf{i})}[u](\mathbf{y})\right|\pi(\mathbf{y})\right)$.

Note that $\pi(\mathbf{y}) > 0$ needs not be equal to $\varrho_{normal}(\mathbf{y})$.

1. $$\Delta W(\mathbf{i}) = \begin{cases} \prod_{n=1}^{N}\left(m(i_n) - m(i_n - 1)\right) & \text{for Genz–Keister} \\ \prod_{n=1}^{N} m(i_n) & \text{for Gauss–Hermite} \end{cases}$$

   or alternatively $\Delta W(\mathbf{i}) = 1$.

2. If using Genz–Keister (nested) knots, the sparse grid is interpolatory, hence

   $$\Delta E(\mathbf{i}) = \max_{\mathbf{y}\in pts(\mathcal{S}_{\mathcal{I}\cup\{\mathbf{i}\}})\setminus pts(\mathcal{S}_{\mathcal{I}})}\left(\left|u(\mathbf{y}) - \mathcal{S}_{\mathcal{I}}^{m}[u](\mathbf{y})\right|\pi(\mathbf{y})\right).$$

   If using Gauss–Hermite (non-nested) knots, the sparse grid is not interpolatory, hence

   $$\Delta E(\mathbf{i}) = \max_{\mathbf{y}\in pts(\mathcal{T}^{m(\mathbf{i})})}\left(\left|\mathcal{S}_{\mathcal{I}\cup\{\mathbf{i}\}}^{m}[u](\mathbf{y}) - \mathcal{S}_{\mathcal{I}}^{m}[u](\mathbf{y})\right|\pi(\mathbf{y})\right).$$

So again, we can use the algorithm with non-nested nodes too

"A-priori" strategy

Method 3: Quasi-optimal sparse grids

more

Here we consider $\Delta E(\mathbf{i}) \approx \left\| \mathbf{\Delta}^{m(\mathbf{i})}[u] \right\|_{L^2_\varrho(\Gamma)}$.

By linking $\mathbf{\Delta}^{m(\mathbf{i})}[u]$ to the spectral (Hermite) expansion of u, **we can get computable estimates of $\Delta E(\mathbf{i})$**, **before** actually running the sparse grid algorithm (see next slide).

$\Delta W(\mathbf{i})$ can also be (easily) estimated (see next slide).

Then, just **compute** the set $\mathcal{I}_k = \left\{ \frac{\Delta E(\mathbf{i})}{\Delta W(\mathbf{i})} > \epsilon_k \right\}$

Method 3: Quasi-optimal sparse grids

more

Here we consider $\Delta E(\mathbf{i}) \approx \left\| \boldsymbol{\Delta}^{m(\mathbf{i})}[u] \right\|_{L^2_\varrho(\Gamma)}$.

Let $\mathcal{H}_{\mathbf{p}}(\mathbf{y})$ be the Hermite pol. of degree p_n wrt y_n.

1. Let $m(i_n) = i_n$ for Gauss–Hermite pts (non-nested), $m(i_n) = 1, 3, 9, 19, 35$ for Genz–Keister (nested). Then let
$$\Delta W(\mathbf{i}) = \begin{cases} \prod_{n=1}^{N} \left(m(i_n) - m(i_n - 1) \right) & \text{Genz–Keister} \\ \prod_{n=1}^{N} m(i_n) & \text{Gauss–Hermite} \end{cases}$$
2. $\Delta E(\mathbf{i}) = C \dfrac{e^{-\sum_{n=1}^{N} g_n m(i_n - 1)}}{\prod_{n=1}^{N} \sqrt{m(i_n - 1)}} B(\mathbf{i})$, with $B(\mathbf{i}) = \left\| \boldsymbol{\Delta}^{m(\mathbf{i})}[\mathcal{H}_{m(\mathbf{i}-\mathbf{1})}] \right\|_{L^2_\varrho(\Gamma)}$
3. For a suff. large "universe" $\mathbb{U} \subset \mathbb{N}^N$ compute the profit of each $\boldsymbol{\Delta}^{m(\mathbf{i})}$, $P(\mathbf{i}) = \Delta E(\mathbf{i}) / \Delta W(\mathbf{i})$
4. sort decreasingly $P(\mathbf{i})$
5. build the sparse grid with the $\boldsymbol{\Delta}^{m(\mathbf{i})}$ with largest profits

In order to compute the profits we need to assess the rates g_n. This can be performed numerically with a comp. cost linear in N ("a priori/a posteriori approach")

Method 3: Quasi-optimal sparse grids

more

Here we consider $\Delta E(\mathbf{i}) \approx \left\| \boldsymbol{\Delta}^{m(\mathbf{i})}[u] \right\|_{L^2_\varrho(\Gamma)}$.

Let $\mathcal{H}_{\mathbf{p}}(\mathbf{y})$ be the Hermite pol. of degree p_n wrt y_n.

1. Let $m(i_n) = i_n$ for Gauss–Hermite pts (non-nested), $m(i_n) = 1, 3, 9, 19, 35$ for Genz–Keister (nested). Then let
$$\Delta W(\mathbf{i}) = \begin{cases} \prod_{n=1}^N \left(m(i_n) - m(i_n - 1)\right) & \text{Genz–Keister} \\ \prod_{n=1}^N m(i_n) & \text{Gauss–Hermite} \end{cases}$$
2. $\Delta E(\mathbf{i}) = C \dfrac{e^{-\sum_{n=1}^N g_n m(i_n - 1)}}{\prod_{n=1}^N \sqrt{m(i_n - 1)}} B(\mathbf{i})$, with $B(\mathbf{i}) = \left\| \boldsymbol{\Delta}^{m(\mathbf{i})}[\mathcal{H}_{m(\mathbf{i}-\mathbf{1})}] \right\|_{L^2_\varrho(\Gamma)}$
3. For a decreasing sequence $\epsilon_1, \epsilon_2 \dots$ compute the sets
$$\mathcal{I}_k = \left\{ \frac{\Delta E(\mathbf{i})}{\Delta W(\mathbf{i})} > \epsilon_k \right\}$$

In order to compute the profits we need to assess the rates g_n. This can be performed numerically with a comp. cost linear in N ("a priori/a posteriori approach")

Method 3: Quasi-optimal sparse grids

more

Here we consider $\Delta E(\mathbf{i}) \approx \left\| \mathbf{\Delta}^{m(\mathbf{i})}[u] \right\|_{L^2_\varrho(\Gamma)}$.

Let $\mathcal{H}_{\mathbf{p}}(\mathbf{y})$ be the Hermite pol. of degree p_n wrt y_n.

1. Let $m(i_n) = i_n$ for Gauss–Hermite pts (non-nested), $m(i_n) = 1, 3, 9, 19, 35$ for Genz–Keister (nested). Then let
$$\Delta W(\mathbf{i}) = \begin{cases} \prod_{n=1}^N \left(m(i_n) - m(i_n - 1) \right) & \text{Genz–Keister} \\ \prod_{n=1}^N m(i_n) & \text{Gauss–Hermite} \end{cases}$$
2. $\Delta E(\mathbf{i}) = C \dfrac{e^{-\sum_{n=1}^N g_n m(i_n - 1)}}{\prod_{n=1}^N \sqrt{m(i_n - 1)}} B(\mathbf{i})$, with $B(\mathbf{i}) = \left\| \mathbf{\Delta}^{m(\mathbf{i})}[\mathcal{H}_{m(\mathbf{i}-\mathbf{1})}] \right\|_{L^2_\varrho(\Gamma)}$
3. For a decreasing sequence $\epsilon_1, \epsilon_2 \dots$ compute the sets
$$\mathcal{I}_k = \left\{ \frac{\frac{e^{-\sum_{n=1}^N g_n m(i_n - 1)}}{\prod_{n=1}^N \sqrt{m(i_n - 1)}} B(\mathbf{i})}{\Delta W} > \epsilon_k \right\}$$

In order to compute the profits we need to assess the rates g_n. This can be performed numerically with a comp. cost linear in N ("a priori/a posteriori approach")

Method 3: Quasi-optimal sparse grids

more

Here we consider $\Delta E(\mathbf{i}) \approx \left\| \mathbf{\Delta}^{m(\mathbf{i})}[u] \right\|_{L^2_\varrho(\Gamma)}$.

Let $\mathcal{H}_{\mathbf{p}}(\mathbf{y})$ be the Hermite pol. of degree p_n wrt y_n.

1. Let $m(i_n) = i_n$ for Gauss–Hermite pts (non-nested), $m(i_n) = 1, 3, 9, 19, 35$ for Genz–Keister (nested). Then let
$$\Delta W(\mathbf{i}) = \begin{cases} \prod_{n=1}^N \left(m(i_n) - m(i_n - 1) \right) & \text{Genz–Keister} \\ \prod_{n=1}^N m(i_n) & \text{Gauss–Hermite} \end{cases}$$
2. $\Delta E(\mathbf{i}) = C \dfrac{e^{-\sum_{n=1}^N g_n m(i_n - 1)}}{\prod_{n=1}^N \sqrt{m(i_n - 1)}} B(\mathbf{i})$, with $B(\mathbf{i}) = \left\| \mathbf{\Delta}^{m(\mathbf{i})}[\mathcal{H}_{m(\mathbf{i}-\mathbf{1})}] \right\|_{L^2_\varrho(\Gamma)}$
3. For a decreasing sequence $\epsilon_1, \epsilon_2 \ldots$ compute the sets
$$\mathcal{I}_k = \left\{ \log \left(\frac{e^{-\sum_{n=1}^N g_n m(i_n - 1)} B(\mathbf{i})}{\prod_{n=1}^N \sqrt{m(i_n - 1)}\, \Delta W} \right) > \log \epsilon_k \right\}$$

In order to compute the profits we need to assess the rates g_n. This can be performed numerically with a comp. cost linear in N ("a priori/a posteriori approach")

Method 3: Quasi-optimal sparse grids

more

Here we consider $\Delta E(\mathbf{i}) \approx \left\| \boldsymbol{\Delta}^{m(\mathbf{i})}[u] \right\|_{L^2_\varrho(\Gamma)}$.

Let $\mathcal{H}_{\mathbf{p}}(\mathbf{y})$ be the Hermite pol. of degree p_n wrt y_n.

1. Let $m(i_n) = i_n$ for Gauss–Hermite pts (non-nested), $m(i_n) = 1, 3, 9, 19, 35$ for Genz–Keister (nested). Then let
$$\Delta W(\mathbf{i}) = \begin{cases} \prod_{n=1}^N \left(m(i_n) - m(i_n - 1)\right) & \text{Genz–Keister} \\ \prod_{n=1}^N m(i_n) & \text{Gauss–Hermite} \end{cases}$$
2. $\Delta E(\mathbf{i}) = C \dfrac{e^{-\sum_{n=1}^N g_n m(i_n - 1)}}{\prod_{n=1}^N \sqrt{m(i_n - 1)}} B(\mathbf{i})$, with $B(\mathbf{i}) = \left\| \boldsymbol{\Delta}^{m(\mathbf{i})}[\mathcal{H}_{m(\mathbf{i}-\mathbf{1})}] \right\|_{L^2_\varrho(\Gamma)}$
3. For an integer sequence $w_1, w_2 \ldots$ compute the sets

$$\mathcal{I}_k = \left\{ \sum_{n=1}^N g_n m(i_n - 1) - \log B(\mathbf{i}) + \log \sum \sqrt{m(i_n - 1)} + \log \Delta W(\mathbf{i}) < w_k \right\}$$

In order to compute the profits we need to assess the rates g_n. This can be performed numerically with a comp. cost linear in N ("a priori/a posteriori approach")

Convergence theorem for quasi-optimal sparse grids

If the profits satisfy the weighted summability condition

$$\left(\sum_{\mathbf{i} \in \mathbb{N}^N} P^\tau(\mathbf{i}) \Delta W(\mathbf{i}) \right)^{1/\tau} = C_P(\tau) < \infty$$

for some $0 < \tau < 1$, then the knapsack sparse grid approximation of u satisfies

$$\| u - \mathcal{S}_{\mathcal{I}}[u] \| \leq W^{1-1/\tau} C_P(\tau).$$

where W is the number of collocation points used to build $\mathcal{S}_{\mathcal{I}}[u]$.

Sketch of proof:

- Let $\{Q_j\}_{j \in \mathbb{N}_+} = \sum_{k=1}^{j} \Delta W_k$ and
$$\{\widetilde{P}_k\}_{k \in \mathbb{N}_+} = \underbrace{\frac{\Delta E_1}{\Delta W_1}, \frac{\Delta E_1}{\Delta W_1}, \frac{\Delta E_1}{\Delta W_1} \ldots}_{\Delta W_1 \textit{ times}}, \underbrace{\frac{\Delta E_2}{\Delta W_2}, \frac{\Delta E_2}{\Delta W_2}, \frac{\Delta E_2}{\Delta W_2} \ldots}_{\Delta W_2 \textit{ times}}.$$
- $\left\| u - \mathcal{S}^m_{\mathcal{I}(w)}[u] \right\| \leq \sum_{\mathbf{i} \notin \mathcal{I}(w)} \left\| \Delta^{m(\mathbf{i})} u \right\| = \sum_{j > w} \Delta E_j = \sum_{k > Q_w} \widetilde{P}_k$.
- Use Stechkin Lemma:
$$\sum_{k > Q_w} \widetilde{P}_k \leq Q_w^{-1/\tau+1} \left(\sum_{k>0} \widetilde{P}_k^\tau \right)^{1/\tau} = Q_w^{-1/\tau+1} \left(\sum_{k>0} P_k^\tau \Delta W_k \right)^{1/\tau}$$

It is quite easy to extend the proof for cases where $\mathcal{I}$ is not admissible.

Convergence theorem for quasi-optimal sparse grids

If the profits satisfy the weighted summability condition

$$\Big(\sum_{\mathbf{i} \in \mathbb{N}^N} P^\tau(\mathbf{i}) \Delta W(\mathbf{i}) \Big)^{1/\tau} = C_P(\tau) < \infty$$

for some $0 < \tau < 1$, then the knapsack sparse grid approximation of u satisfies

$$\| u - \mathcal{S}_\mathcal{I}[u] \| \leq W^{1-1/\tau} C_P(\tau).$$

where W is the number of collocation points used to build $\mathcal{S}_\mathcal{I}[u]$.

Ex.: assume for a moment $y_1, \ldots, y_N$ uniform random variables, sparse grids with Clenshaw–Curtis points, Legendre coeff. decay as $C(N) e^{-\sum_{n=1}^N g_n m(i_n - 1)}$. Then the theorem above implies that

$$\| u - \mathcal{S}_\mathcal{I}[u] \| \leq C_1(N) \exp\left(-C_2 N \sqrt[N]{W} \right)$$

Sketch of proof: observe that in this case P are τ-summable for $0 < \tau < 1$; then, optimize the error estimate w.r.t. τ.

To refine or to add random var.? Dimension-adaptivity

- **Problem:** exploring reduced margin in high-dimensional spaces / computing all rates g_n beforehand is too expensive.

To refine or to add random var.? Dimension-adaptivity

- **Problem:** exploring reduced margin in high-dimensional spaces / computing all rates g_n beforehand is too expensive.
- **Assume** that the Karhunen–Loève expansion $\gamma = \sigma \sum_{k=1}^{N} y_k \gamma_k \phi_k$ introduces a "weak ordering" of random variables, i.e. there exists $B \geq 1$ (buffer) s.t. y_{n+B} is guaranteed to be less important than y_n.

To refine or to add random var.? Dimension-adaptivity

- **Problem:** exploring reduced margin in high-dimensional spaces / computing all rates g_n beforehand is too expensive.
- **Assume** that the Karhunen–Loève expansion $\gamma = \sigma \sum_{k=1}^{N} y_k \gamma_k \phi_k$ introduces a "weak ordering" of random variables, i.e. there exists $B \geq 1$ (buffer) s.t. y_{n+B} is guaranteed to be less important than y_n.
- **Idea:** Add random variables **gradually** (balance refinement and addition of variables). Note that this means we don't need to truncate a-priori the Karhunen–Loève expansion of γ!

To refine or to add random var.? Dimension-adaptivity

- **Problem:** exploring reduced margin in high-dimensional spaces / computing all rates g_n beforehand is too expensive.
- **Assume** that the Karhunen–Loève expansion $\gamma = \sigma \sum_{k=1}^{N} y_k \gamma_k \phi_k$ introduces a "weak ordering" of random variables, i.e. there exists $B \geq 1$ (buffer) s.t. y_{n+B} is guaranteed to be less important than y_n.
- **Idea:** Add random variables **gradually** (balance refinement and addition of variables). Note that this means we don't need to truncate a-priori the Karhunen–Loève expansion of γ!
- **Define** a random variable y_n as **activated** if $\min_{\mathbf{i} \in \mathcal{I}} i_n > 1$.

To refine or to add random var.? Dimension-adaptivity

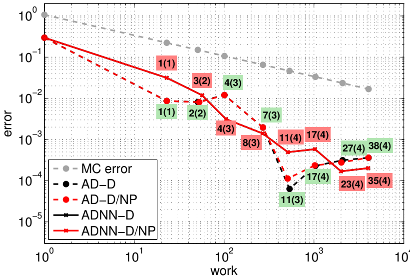
- **Problem:** exploring reduced margin in high-dimensional spaces / computing all rates g_n beforehand is too expensive.
- **Assume** that the Karhunen–Loève expansion $\gamma = \sigma \sum_{k=1}^{N} y_k \gamma_k \phi_k$ introduces a "weak ordering" of random variables, i.e. there exists $B \geq 1$ (buffer) s.t. y_{n+B} is guaranteed to be less important than y_n.
- **Idea:** Add random variables **gradually** (balance refinement and addition of variables). Note that this means we don't need to truncate a-priori the Karhunen–Loève expansion of γ!
- **Define** a random variable y_n as **activated** if $\min_{\mathbf{i} \in \mathcal{I}} i_n > 1$.

A simple dimension-adaptive algorithm

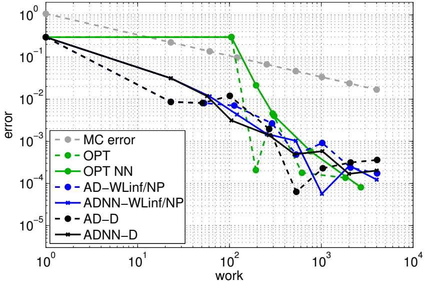
1. start the adaptive/quasi-optimal algorithm using B rand. var.
2. As soon as one of these "buffer variables" gets activated, add a new rand. var. to the approximaton (in the quasi optimal setting, this means computing the corresponding g_n).

The uncertain Darcy problem – results 1

Field data: $\sigma = 1$, corr. length $L_c = 0.5$, $\nu = 2.5$ (C^2 realizations).



adaptive schemes with activated variables



adaptive and quasi-optimal schemes

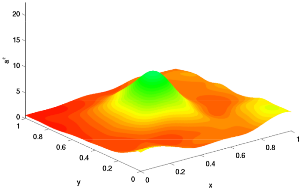
- moderate number of rand. vars. needed
- quasi-optimal plateau due to computation of g
- quasi-optimal and adaptive schemes have similar convergence
- nested and non-nested points have similar convergence
- convergence robust wrt. type of points and $\mathbb{E}[\cdot]/L^\infty$-driven adaptation
- similar results for adaptive sparse interpolation

Case $\nu = 0.5$: Monte Carlo Control Variate

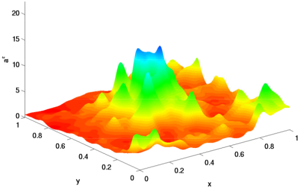
Realizations of γ are non-differentiable $\Rightarrow$ KL eigenval decay very slow $\Rightarrow$ sparse grids may be non-effective.

Remedy: use sparse grids as **control var.** (preconditioner) for MC

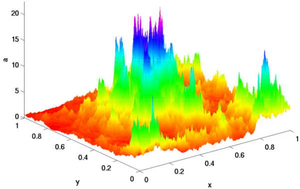
1. Consider a **smoothed field** γ^ϵ, such that $\mathcal{Q}_\mathcal{I}[u^\epsilon] \to \mathbb{E}[u^\epsilon]$ quickly.



smoothed field, $\epsilon = 1/2^4$



smoothed field $\epsilon = 1/2^6$



non-smoothed field, $\epsilon = 0$

Case $\nu = 0.5$: Monte Carlo Control Variate

Realizations of γ are non-differentiable $\Rightarrow$ KL eigenval decay very slow $\Rightarrow$ sparse grids may be non-effective.

Remedy: use sparse grids as **control var.** (preconditioner) for MC

1. Consider a **smoothed field** γ^ϵ, such that $\mathcal{Q}_\mathcal{I}[u^\epsilon] \to \mathbb{E}[u^\epsilon]$ quickly.
2. Define $u_{CV} = u - u^\epsilon + \mathcal{Q}_\mathcal{I}[u^\epsilon]$. There holds

$$\mathbb{E}[u_{CV}] = \mathbb{E}[u], \quad \mathbb{V}\mathrm{ar}(u_{CV}) = \mathbb{V}\mathrm{ar}(u) + \mathbb{V}\mathrm{ar}(u^\epsilon) - 2\mathrm{cov}(u, u^\epsilon)$$

Thus, the smaller ϵ, the smaller the MC error, but slower the convergence $\mathcal{Q}_\mathcal{I}[u^\epsilon] \to \mathbb{E}[u^\epsilon]$.

Case $\nu = 0.5$: Monte Carlo Control Variate

Realizations of γ are non-differentiable $\Rightarrow$ KL eigenval decay very slow $\Rightarrow$ sparse grids may be non-effective.

Remedy: use sparse grids as **control var.** (preconditioner) for MC

1. Consider a **smoothed field** γ^ϵ, such that $\mathcal{Q}_\mathcal{I}[u^\epsilon] \to \mathbb{E}[u^\epsilon]$ quickly.
2. Define $u_{CV} = u - u^\epsilon + \mathcal{Q}_\mathcal{I}[u^\epsilon]$. There holds

$$\mathbb{E}[u_{CV}] = \mathbb{E}[u], \quad \mathbb{V}\mathrm{ar}(u_{CV}) = \mathbb{V}\mathrm{ar}(u) + \mathbb{V}\mathrm{ar}(u^\epsilon) - 2\mathrm{cov}(u, u^\epsilon)$$

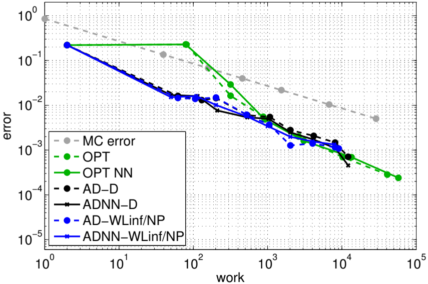
   Thus, the smaller ϵ, the smaller the MC error, but slower the convergence $\mathcal{Q}_\mathcal{I}[u^\epsilon] \to \mathbb{E}[u^\epsilon]$.

3. Set $\mathbb{E}[u_{CV}] \approx \frac{1}{M}\sum_{i=1}^{M} u^{CV}(\omega_i) = \frac{1}{M}\sum_{i=1}^{M} (u(\omega_i) - u^\epsilon(\omega_i)) + \mathcal{Q}_\mathcal{I}^m[u^\epsilon]$.

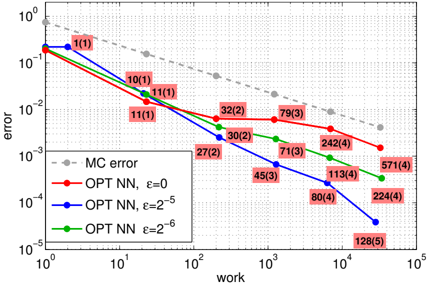
   M can be chosen balancing either the works or the errors of MC and sparse grids.

The uncertain Darcy problem – results 2

Field data: $\sigma = 1$, corr. length $L_c = 0.5$, $\nu = 0.5$



MCCV error for adaptive and quasi-optimal sparse grids. ~ 30 r.v. activated.



Sparse grid component of the error for different values of ϵ. The performance deteriorates as $\epsilon \to 0$

Conclusions

1. General framework for quasi-optimal and adaptive sparse grids schemes;
2. The schemes can be applied to the lognormal case, also with non-nested points;
3. The dimension-adaptive implementation allows to work without a-priori truncation of the random field;
4. A Monte Carlo Control Variate can be used to improve results in the rough case (exponential covariance).

Thank you for your attention!

(see also poster)

Bibliography

- F. Nobile, L. Tamellini, F. Tesei and R. Tempone. An adaptive sparse grid algorithm for elliptic PDEs with lognormal diffusion coefficient Sparse Grids and Applications 2014, Springer (to appear). Also available as MATHICSE report 4/2015
- F. Tesei Numerical Approximation of Flows in Random Porous Media Ph.D. Thesis Nr. 6860, EPFL, to appear on http://infoscience.epfl.ch
- F. Tesei and F. Nobile. A Multi Level Monte Carlo Method with Control Variate for elliptic PDEs with log-normal coefficients. Stochastic Partial Differential Equations: Analysis and Computations, 3(3), 2015.
- F. Nobile, L. Tamellini, and R. Tempone. Convergence of quasi-optimal sparse-grid approximation of Hilbert-space–valued functions: application to random elliptic PDEs. Numerische Mathematik (available online)
- J. Beck, F. Nobile, L. Tamellini, and R. Tempone. A Quasi-optimal Sparse Grids Procedure for Groundwater Flows. Selected papers from the ICOSAHOM '12 conference.
- J. Beck, F. Nobile, L. Tamellini, and R. Tempone. On the optimal polynomial approximation of stochastic PDEs by Galerkin and Collocation methods Math. Mod. Methods Appl. Sci. (M3AS) 22(9), 2012

Matérn covariance

$$cov_{\gamma_{\nu}}(\mathbf{x},\mathbf{x}') = \sigma^2 \prod_{i=1}^{2} \frac{\left(\sqrt{2\nu}|x_i - x_i'|/L_c\right)^{\nu} K_{\nu}\left(\sqrt{2\nu}|x_i - x_i'|/L_c\right)}{G(\nu)2^{\nu-1}}$$

Here K_{ν} is the modified Bessel function of the second kind and G the Gamma function. back

Knapsack approach

$\mathcal{S}_{\mathcal{I}}[u](\mathbf{y}) = \sum_{\mathbf{i} \in \mathcal{I}} \mathbf{\Delta}^{m(\mathbf{i})}[u](\mathbf{y})$. **What terms $\mathbf{\Delta}^{m(\mathbf{i})}$ to include in the sum?**

Given that $\mathcal{E}\left[u - \mathcal{S}_{\mathcal{I}}[u]\right] = \mathcal{E}\left[\sum_{\mathbf{i} \notin \mathcal{I}} \mathbf{\Delta}^{m(\mathbf{i})}[u]\right] \leq \sum_{\mathbf{i} \notin \mathcal{I}} \mathcal{E}\left[\mathbf{\Delta}^{m(\mathbf{i})}[u]\right],$

Knapsack approach

$\mathcal{S}_{\mathcal{I}}[u](\mathbf{y}) = \sum_{\mathbf{i}\in\mathcal{I}} \mathbf{\Delta}^{m(\mathbf{i})}[u](\mathbf{y})$. **What terms $\mathbf{\Delta}^{m(\mathbf{i})}$ to include in the sum?**

Given that $\mathcal{E}\left[u - \mathcal{S}_{\mathcal{I}}[u]\right] = \mathcal{E}\left[\sum_{\mathbf{i}\notin\mathcal{I}} \mathbf{\Delta}^{m(\mathbf{i})}[u]\right] \leq \sum_{\mathbf{i}\notin\mathcal{I}} \Delta E(\mathbf{i})$

1 – Set

- $\mathcal{E}\left[\mathbf{\Delta}^{m(\mathbf{i})}[u]\right] \approx \Delta E(\mathbf{i})$
- $\Delta W(\mathbf{i})$ = work of $\mathbf{\Delta}^{m(\mathbf{i})}[u]$
- $x_{\mathbf{i}} \in \{0,1\}$, $x_{\mathbf{i}} = 1 \Leftrightarrow \mathbf{i} \in \mathcal{I}$.

Knapsack approach

$\mathcal{S}_{\mathcal{I}}[u](\mathbf{y}) = \sum_{\mathbf{i} \in \mathcal{I}} \mathbf{\Delta}^{m(\mathbf{i})}[u](\mathbf{y})$. **What terms $\mathbf{\Delta}^{m(\mathbf{i})}$ to include in the sum?**

Given that $\mathcal{E}\left[u - \mathcal{S}_{\mathcal{I}}[u]\right] = \mathcal{E}\left[\sum_{\mathbf{i} \notin \mathcal{I}} \mathbf{\Delta}^{m(\mathbf{i})}[u]\right] \leq \sum_{\mathbf{i} \notin \mathcal{I}} \Delta E(\mathbf{i})$

1 – Set

- $\mathcal{E}\left[\mathbf{\Delta}^{m(\mathbf{i})}[u]\right] \approx \Delta E(\mathbf{i})$
- $\Delta W(\mathbf{i}) =$ work of $\mathbf{\Delta}^{m(\mathbf{i})}[u]$
- $x_{\mathbf{i}} \in \{0,1\}$, $x_{\mathbf{i}} = 1 \Leftrightarrow \mathbf{i} \in \mathcal{I}$.

2 – Consider

$\max \sum_{\mathbf{i} \in \mathbb{N}_+^N} \Delta E(\mathbf{i}) x_{\mathbf{i}}$ s.t.

- $\sum_{\mathbf{i} \in \mathbb{N}_+^N} \Delta W(\mathbf{i}) x_{\mathbf{i}} \leq W_{max}$ and
- $x_{\mathbf{i}} \in \{0,1\}$

Knapsack approach

$\mathcal{S}_{\mathcal{I}}[u](\mathbf{y}) = \sum_{\mathbf{i}\in\mathcal{I}} \boldsymbol{\Delta}^{m(\mathbf{i})}[u](\mathbf{y})$. **What terms $\boldsymbol{\Delta}^{m(\mathbf{i})}$ to include in the sum?**

Given that $\mathcal{E}\left[u - \mathcal{S}_{\mathcal{I}}[u]\right] = \mathcal{E}\left[\sum_{\mathbf{i}\notin\mathcal{I}} \boldsymbol{\Delta}^{m(\mathbf{i})}[u]\right] \leq \sum_{\mathbf{i}\notin\mathcal{I}} \Delta E(\mathbf{i})$

1 – Set

- $\mathcal{E}\left[\boldsymbol{\Delta}^{m(\mathbf{i})}[u]\right] \approx \Delta E(\mathbf{i})$
- $\Delta W(\mathbf{i})$ = work of $\boldsymbol{\Delta}^{m(\mathbf{i})}[u]$
- $x_{\mathbf{i}} \in \{0,1\}$, $x_{\mathbf{i}} = 1 \Leftrightarrow \mathbf{i} \in \mathcal{I}$.

2 – Consider

$\max \sum_{\mathbf{i}\in\mathbb{N}_+^N} \Delta E(\mathbf{i}) x_{\mathbf{i}}$ s.t.

- $\sum_{\mathbf{i}\in\mathbb{N}_+^N} \Delta W(\mathbf{i}) x_{\mathbf{i}} \leq W_{max}$ and
- $x_{\mathbf{i}} \in \{0,1\}$

3 – Solve by Dantzig Algorithm (knapsack problem approach)

1. compute the profit of each $\boldsymbol{\Delta}^{m(\mathbf{i})}[u]$, $P(\mathbf{i}) = \Delta E(\mathbf{i}) / \Delta W(\mathbf{i})$
2. sort $\boldsymbol{\Delta}^{m(\mathbf{i})}[u]$ decreasing in profits
3. set $x_{\mathbf{i}} = 1$ until W_{max} is reached

back

Quasi-optimal sparse grids

back

$$\Delta E(\mathbf{i}) = \left\| \mathbf{\Delta}^{m(\mathbf{i})}[u] \right\| = \left\| \mathbf{\Delta}^{m(\mathbf{i})} \Big[\sum_{\mathbf{q} \in \mathbb{N}^N} u_{\mathbf{q}} \mathcal{H}_{\mathbf{q}} \Big] \right\| = \left\| \sum_{\mathbf{q} \in \mathbb{N}^N} u_{\mathbf{q}} \mathbf{\Delta}^{m(\mathbf{i})}[\mathcal{H}_{\mathbf{q}}] \right\|$$

Next, by construction $\mathbf{\Delta}^{m(\mathbf{i})}[\mathcal{H}_{\mathbf{q}}] = 0$ for polynomials such that $\exists\, n : q_n < m(i_n - 1)$. By triangular inequality we get

$$\Delta E(\mathbf{i}) \leq \sum_{\mathbf{q} \geq m(\mathbf{i}-\mathbf{1})} \|u_{\mathbf{q}}\|_{H^1(D)} \left\| \mathbf{\Delta}^{m(\mathbf{i})}[\mathcal{H}_{\mathbf{q}}] \right\|_{L^2_\rho(\Gamma)} .$$

By assuming that the summation is dominated by the first term, we get

$$\Delta E(\mathbf{i}) \approx B(\mathbf{i}) \left\| u_{m(\mathbf{i}-\mathbf{1})} \right\|_{H^1(D)}, \quad B(\mathbf{i}) = \left\| \mathbf{\Delta}^{m(\mathbf{i})}[\mathcal{H}_{m(\mathbf{i}-\mathbf{1})}] \right\|_{L^2_\rho(\Gamma)}$$

and we use the Heuristic $\left\| u_{m(\mathbf{i}-\mathbf{1})} \right\|_{H^1(D)} \leq \dfrac{e^{-\sum_{n=1}^N g_n m(i_n - 1)}}{\prod_{n=1}^N \sqrt{m(i_n - 1)}}$