Should we be Greedy?

Ronald DeVore

Texas A& M University - UPMC

The Beginnings

- This story begins with Yvon Maday

The Beginnings

- This story begins with Yvon Maday
- Yvon explained a certain greedy algorithm used to find what is called a reduced basis for simultaneously solving a family of parametric PDEs

The Beginnings

- This story begins with Yvon Maday
- Yvon explained a certain greedy algorithm used to find what is called a reduced basis for simultaneously solving a family of parametric PDEs
- This algorithm had not been studied by the approximation theory community and the results of Yvon and his collaborators were at first glance somewhat puzzling to us

The Beginnings

- This story begins with Yvon Maday
- Yvon explained a certain greedy algorithm used to find what is called a reduced basis for simultaneously solving a family of parametric PDEs
- This algorithm had not been studied by the approximation theory community and the results of Yvon and his collaborators were at first glance somewhat puzzling to us
- This talk is an attempt to understand this algorithm and how it performs

The Approximation Challenge

- All action will take place in a Hilbert space $\mathcal{H}$

The Approximation Challenge

- All action will take place in a Hilbert space $\mathcal{H}$
- We are given a compact set of functions $\mathcal{K}$ which we wish to approximate

The Approximation Challenge

- All action will take place in a Hilbert space $\mathcal{H}$
- We are given a compact set of functions $\mathcal{K}$ which we wish to approximate
- We look for a finite number of functions $f_0, \ldots, f_{N-1} \in \mathcal{K}$ such that for $V_N := \text{span}\{f_0, \ldots, f_{N-1}\}$, we have

$$\text{dist}(f, V_N) \leq \epsilon, \quad \forall f \in \mathcal{K}$$

The Approximation Challenge

- All action will take place in a Hilbert space $\mathcal{H}$
- We are given a compact set of functions $\mathcal{K}$ which we wish to approximate
- We look for a finite number of functions $f_0, \ldots, f_{N-1} \in \mathcal{K}$ such that for $V_N := \text{span}\{f_0, \ldots, f_{N-1}\}$, we have

$$\text{dist}(f, V_N) \leq \epsilon, \quad \forall f \in \mathcal{K}$$

- Of course for the given tolerance ϵ we want N to be as small as possible

The (Pure) Greedy Algorithm

- $f_0 := \text{argmax}\{\|f\| : f \in \mathcal{K}\}$

The (Pure) Greedy Algorithm

- $f_0 := \operatorname{argmax}\{\|f\| : f \in \mathcal{K}\}$
- If $f_0, \ldots, f_{n-1}$ have been chosen, define

The (Pure) Greedy Algorithm

- $f_0 := \text{argmax}\{\|f\| : f \in \mathcal{K}\}$
- If $f_0, \ldots, f_{n-1}$ have been chosen, define
  - $V_n := \text{span}\{f_0, \ldots, f_{n-1}\}$

The (Pure) Greedy Algorithm

- $f_0 := \operatorname{argmax}\{\|f\| : f \in \mathcal{K}\}$
- If $f_0, \ldots, f_{n-1}$ have been chosen, define
  - $V_n := \operatorname{span}\{f_0, \ldots, f_{n-1}\}$
  - P_{V_n} as the orthogonal projector onto V_n

The (Pure) Greedy Algorithm

- $f_0 := \operatorname{argmax}\{\|f\| : f \in \mathcal{K}\}$
- If $f_0, \dots, f_{n-1}$ have been chosen, define
  - $V_n := \operatorname{span}\{f_0, \dots, f_{n-1}\}$
  - P_{V_n} as the orthogonal projector onto V_n
  - $f_n := \operatorname*{Argmax}_{f \in \mathcal{K}} \|f - P_{V_n} f\|_{\mathcal{H}}$

The (Pure) Greedy Algorithm

- $f_0 := \text{argmax}\{\|f\| : f \in \mathcal{K}\}$
- If $f_0, \ldots, f_{n-1}$ have been chosen, define
  - $V_n := \text{span}\{f_0, \ldots, f_{n-1}\}$
  - P_{V_n} as the orthogonal projector onto V_n
  - $f_n := \underset{f \in \mathcal{K}}{\text{Argmax}} \|f - P_{V_n} f\|_{\mathcal{H}}$
- Thus at each step, the function f_n is chosen in a greedy manner

The (Pure) Greedy Algorithm

- $f_0 := \operatorname{argmax}\{\|f\| : f \in \mathcal{K}\}$
- If $f_0, \dots, f_{n-1}$ have been chosen, define
  - $V_n := \operatorname{span}\{f_0, \dots, f_{n-1}\}$
  - P_{V_n} as the orthogonal projector onto V_n
  - $f_n := \operatorname*{Argmax}_{f \in \mathcal{K}} \|f - P_{V_n} f\|_{\mathcal{H}}$
- Thus at each step, the function f_n is chosen in a greedy manner
- Our experience tells us that such greedy strategies are usually not very good

The (Pure) Greedy Algorithm

- $f_0 := \operatorname{argmax}\{\|f\| : f \in \mathcal{K}\}$
- If $f_0, \dots, f_{n-1}$ have been chosen, define
  - $V_n := \operatorname{span}\{f_0, \dots, f_{n-1}\}$
  - P_{V_n} as the orthogonal projector onto V_n
  - $f_n := \operatorname*{Argmax}_{f \in \mathcal{K}} \|f - P_{V_n} f\|_{\mathcal{H}}$
- Thus at each step, the function f_n is chosen in a greedy manner
- Our experience tells us that such greedy strategies are usually not very good
- But before going into an analysis of this algorithm let us point out that as it stands, the algorithm is too idealized for applications as the following example will show

A Simple PDE Setting

- We want to solve a family of elliptic problems

$$\begin{aligned} -\operatorname{div}(a(x,y)\nabla_x u(x,y)) &= f(x), \quad x \in D, \\ u(x,y) &= 0, \quad x \in \partial D \end{aligned}$$

on a domain $D \subset \mathbb{R}^d$ with parameters $y \in U$

A Simple PDE Setting

- We want to solve a family of elliptic problems

$$\begin{aligned} -\operatorname{div}(a(x,y)\nabla_x u(x,y)) &= f(x), \quad x \in D, \\ u(x,y) &= 0, \quad x \in \partial D \end{aligned}$$

on a domain $D \subset \mathbb{R}^d$ with parameters $y \in U$

- We assume uniform ellipticity $0 < r \leq a(\cdot,\cdot) \leq R$

A Simple PDE Setting

- We want to solve a family of elliptic problems

$$\begin{aligned} -\operatorname{div}(a(x,y)\nabla_x u(x,y)) &= f(x), \quad x \in D, \\ u(x,y) &= 0, \quad x \in \partial D \end{aligned}$$

on a domain $D \subset \mathbb{R}^d$ with parameters $y \in U$

- We assume uniform ellipticity $0 < r \leq a(\cdot,\cdot) \leq R$
- The Reduced Basis Method looks for a small set of parameter values $y_1, \ldots, y_n$ such that the $V_n := \operatorname{span}\{u(y_i) : \ 1 \leq i \leq n\}$ is a good Galerkin space:

$$\sup_{y \in U} \|u(y) - P_{V_n} u(y)\|_{H_0^1(a(y))}$$

should be small

A Simple PDE Setting

- We want to solve a family of elliptic problems

$$\begin{aligned} -\operatorname{div}(a(x,y)\nabla_x u(x,y)) &= f(x), \quad x \in D, \\ u(x,y) &= 0, \quad x \in \partial D \end{aligned}$$

on a domain $D \subset \mathbb{R}^d$ with parameters $y \in U$

- We assume uniform ellipticity $0 < r \leq a(\cdot,\cdot) \leq R$
- The Reduced Basis Method looks for a small set of parameter values $y_1, \dots, y_n$ such that the $V_n := \operatorname{span}\{u(y_i) : \ 1 \leq i \leq n\}$ is a good Galerkin space:

$$\sup_{y \in U} \|u(y) - P_{V_n} u(y)\|_{H_0^1(a(y))}$$

should be small

- $H_0^1(a(y))$ is the corresponding energy space

Deficiencies of Pure Greedy

- The greedy algorithm provides a possible strategy for finding a space V_n but there are several issues that prevent its direct use:

Deficiencies of Pure Greedy

- The greedy algorithm provides a possible strategy for finding a space V_n but there are several issues that prevent its direct use:
- Changing energy space:

Deficiencies of Pure Greedy

- The greedy algorithm provides a possible strategy for finding a space V_n but there are several issues that prevent its direct use:
- Changing energy space:
  - The Hilbert spaces $H_0^1(D, a(\cdot, y))$ change with $y \in U$

Deficiencies of Pure Greedy

- The greedy algorithm provides a possible strategy for finding a space V_n but there are several issues that prevent its direct use:
- Changing energy space:
  - The Hilbert spaces $H_0^1(D, a(\cdot, y))$ change with $y \in U$
  - Fortunately, their norms are all equivalent to

$$\|v\|_V := \|v\|_{H_0^1}, \quad \text{i.e. } a \equiv 1$$

Deficiencies of Pure Greedy

- The greedy algorithm provides a possible strategy for finding a space V_n but there are several issues that prevent its direct use:
- Changing energy space:
  - The Hilbert spaces $H_0^1(D, a(\cdot, y))$ change with $y \in U$
  - Fortunately, their norms are all equivalent to

$$\|v\|_V := \|v\|_{H_0^1}, \quad \text{i.e. } a \equiv 1$$

  - So we can use $\mathcal{H} := \mathcal{H}_0^1$ but now we have only equivalent norms

Deficiencies (continued)

- We do not know $u(y)$ so we cant directly compute $\|u(y) - P_{V_n}u(y)\|_{\mathcal{H}}$, for each $y \in U$

Deficiencies (continued)

- We do not know $u(y)$ so we cant directly compute $\|u(y) - P_{V_n} u(y)\|_{\mathcal{H}}$, for each $y \in U$
  - Instead we use surrogate computationally friendly bound $E_n(u(y))$ based on residual estimators

$$C_1 E_n(u(y)) \le \|u(y) - P_{V_n} u(y)\|_{\mathcal{H}} \le C_2 E_n(u(y))$$

Deficiencies (continued)

- We do not know $u(y)$ so we cant directly compute $\|u(y) - P_{V_n}u(y)\|_{\mathcal{H}}$, for each $y \in U$
  - Instead we use surrogate computationally friendly bound $E_n(u(y)$ based on residual estimators

$$C_1 E_n(u(y)) \leq \|u(y) - P_{V_n}u(y)\|_{\mathcal{H}} \leq C_2 E_n(u(y))$$

- Given y, we cant compute $u(y)$ only an approximation $\bar{u}(y)$ - usually done with FEMS

Deficiencies (continued)

- We do not know $u(y)$ so we cant directly compute $\|u(y) - P_{V_n}u(y)\|_{\mathcal{H}}$, for each $y \in U$
  - Instead we use surrogate computationally friendly bound $E_n(u(y)$ based on residual estimators

$$C_1 E_n(u(y)) \leq \|u(y) - P_{V_n}u(y)\|_{\mathcal{H}} \leq C_2 E_n(u(y))$$

- Given y, we cant compute $u(y)$ only an approximation $\bar{u}(y)$ - usually done with FEMS
- Of course we are also going to have to discretize the set $y \in U$

Deficiencies (continued)

- We do not know $u(y)$ so we cant directly compute $\|u(y) - P_{V_n}u(y)\|_{\mathcal{H}}$, for each $y \in U$
  - Instead we use surrogate computationally friendly bound $E_n(u(y))$ based on residual estimators

$$C_1 E_n(u(y)) \leq \|u(y) - P_{V_n}u(y)\|_{\mathcal{H}} \leq C_2 E_n(u(y))$$

- Given y, we cant compute $u(y)$ only an approximation $\bar{u}(y)$ - usually done with FEMS
- Of course we are also going to have to discretize the set $y \in U$
- To handle some of these issues, we consider the following weak greedy algorithm

Weak Greedy Algorithms

- Let $0 < \gamma \leq 1$

Weak Greedy Algorithms

- Let $0 < \gamma \leq 1$
- The choice $\gamma = 1$ gives the ideal greedy algorithm just described

Weak Greedy Algorithms

- Let $0 < \gamma \leq 1$
- The choice $\gamma = 1$ gives the ideal greedy algorithm just described
- The weak greedy algorithm does the following

Weak Greedy Algorithms

- Let $0 < \gamma \leq 1$
- The choice $\gamma = 1$ gives the ideal greedy algorithm just described
- The weak greedy algorithm does the following
  - Choose f_0 such that $\|f_0\| \geq \gamma \max\{\|f\| : f \in \mathcal{K}\}$

Weak Greedy Algorithms

- Let $0 < \gamma \le 1$
- The choice $\gamma = 1$ gives the ideal greedy algorithm just described
- The weak greedy algorithm does the following
  - Choose f_0 such that $\|f_0\| \ge \gamma \max\{\|f\| : f \in \mathcal{K}\}$
  - If $f_0, \dots, f_{n-1}$ have been chosen, define $V_n := \mathrm{span}\{f_0, \dots, f_{n-1}\}$ and P_{V_n}

Weak Greedy Algorithms

- Let $0 < \gamma \leq 1$
- The choice $\gamma = 1$ gives the ideal greedy algorithm just described
- The weak greedy algorithm does the following
  - Choose f_0 such that $\|f_0\| \geq \gamma \max\{\|f\| : f \in \mathcal{K}\}$
  - If $f_0, \ldots, f_{n-1}$ have been chosen, define $V_n := \text{span}\{f_0, \ldots, f_{n-1}\}$ and P_{V_n}
  - Choose f_n so that

$$\|f_n - P_{V_n} f_n\|_{\mathcal{H}} \geq \gamma \max_{f \in \mathcal{K}} \|f - P_{V_n}(f)\|_{\mathcal{H}}$$

Weak Greedy Algorithms

- Let $0 < \gamma \leq 1$
- The choice $\gamma = 1$ gives the ideal greedy algorithm just described
- The weak greedy algorithm does the following
  - Choose f_0 such that $\|f_0\| \geq \gamma \max\{\|f\| : f \in \mathcal{K}\}$
  - If $f_0, \dots, f_{n-1}$ have been chosen, define $V_n := \text{span}\{f_0, \dots, f_{n-1}\}$ and P_{V_n}
  - Choose f_n so that

$$\|f_n - P_{V_n} f_n\|_{\mathcal{H}} \geq \gamma \max_{f \in \mathcal{K}} \|f - P_{V_n}(f)\|_{\mathcal{H}}$$

- Notice that the sequence $f_0, f_1, \dots$ is not unique

Weak Greedy Algorithms

- Let $0 < \gamma \leq 1$
- The choice $\gamma = 1$ gives the ideal greedy algorithm just described
- The weak greedy algorithm does the following
  - Choose f_0 such that $\|f_0\| \geq \gamma \max\{\|f\| : f \in \mathcal{K}\}$
  - If $f_0, \dots, f_{n-1}$ have been chosen, define $V_n := \operatorname{span}\{f_0, \dots, f_{n-1}\}$ and P_{V_n}
  - Choose f_n so that

$$\|f_n - P_{V_n} f_n\|_{\mathcal{H}} \geq \gamma \max_{f \in \mathcal{K}} \|f - P_{V_n}(f)\|_{\mathcal{H}}$$

- Notice that the sequence $f_0, f_1, \dots$ is not unique
- The discussion and analysis below applies to any realization of this algorithm

Performance of Greedy Algorithms

- How should we measure performance of (weak) Greedy Algorithm: compare it with best n dimensional space

Performance of Greedy Algorithms

- How should we measure performance of (weak) Greedy Algorithm: compare it with best n dimensional space
- Kolmogorov width: $d_n(\mathcal{K}) := \inf_{\dim(V)=n} \sup_{f\in\mathcal{K}} \operatorname{dist}(f, V)$

Performance of Greedy Algorithms

- How should we measure performance of (weak) Greedy Algorithm: compare it with best n dimensional space
- Kolmogorov width: $d_n(\mathcal{K}) := \inf_{\dim(V)=n} \sup_{f \in \mathcal{K}} \operatorname{dist}(f, V)$
- This is a benchmark for performance

Performance of Greedy Algorithms

- How should we measure performance of (weak) Greedy Algorithm: compare it with best n dimensional space
- Kolmogorov width: $d_n(\mathcal{K}) := \inf_{\dim(V)=n} \sup_{f\in\mathcal{K}} \mathrm{dist}(f,V)$
- This is a benchmark for performance
- Let $V_n := \mathrm{span}\{f_0,\dots,f_{n-1}\}$ be the subspace generated by the weak greedy selection

Performance of Greedy Algorithms

- How should we measure performance of (weak) Greedy Algorithm: compare it with best n dimensional space
- Kolmogorov width: $d_n(\mathcal{K}) := \inf_{\dim(V)=n} \sup_{f\in\mathcal{K}} \operatorname{dist}(f, V)$
- This is a benchmark for performance
- Let $V_n := \operatorname{span}\{f_0, \dots, f_{n-1}\}$ be the subspace generated by the weak greedy selection
- $\sigma_n(\mathcal{K}) := \sup_{f\in\mathcal{K}} \operatorname{dist}(f, V_n)$ algorithm performance

Performance of Greedy Algorithms

- How should we measure performance of (weak) Greedy Algorithm: compare it with best n dimensional space
- Kolmogorov width: $d_n(\mathcal{K}) := \inf_{\dim(V)=n} \sup_{f\in\mathcal{K}} \operatorname{dist}(f,V)$
- This is a benchmark for performance
- Let $V_n := \operatorname{span}\{f_0,\dots,f_{n-1}\}$ be the subspace generated by the weak greedy selection
- $\sigma_n(\mathcal{K}) := \sup_{f\in\mathcal{K}} \operatorname{dist}(f,V_n)$ algorithm performance
- Buffa, Maday, Patera, Prud'home, Turinici (BMPPT)

Performance of Greedy Algorithms

- How should we measure performance of (weak) Greedy Algorithm: compare it with best n dimensional space
- Kolmogorov width: $d_n(\mathcal{K}) := \inf_{\dim(V)=n} \sup_{f\in\mathcal{K}} \mathrm{dist}(f, V)$
- This is a benchmark for performance
- Let $V_n := \mathrm{span}\{f_0, \dots, f_{n-1}\}$ be the subspace generated by the weak greedy selection
- $\sigma_n(\mathcal{K}) := \sup_{f\in\mathcal{K}} \mathrm{dist}(f, V_n)$ algorithm performance
- Buffa, Maday, Patera, Prud'home, Turinici (BMPPT)
- BMPPT prove that for the ideal greedy $\sigma_n(\mathcal{K}) \le Cn2^n d_n(\mathcal{K})$

Performance of Greedy Algorithms

- How should we measure performance of (weak) Greedy Algorithm: compare it with best n dimensional space
- Kolmogorov width: $d_n(\mathcal{K}) := \inf_{\dim(V)=n} \sup_{f\in\mathcal{K}} \mathrm{dist}(f,V)$
- This is a benchmark for performance
- Let $V_n := \mathrm{span}\{f_0,\dots,f_{n-1}\}$ be the subspace generated by the weak greedy selection
- $\sigma_n(\mathcal{K}) := \sup_{f\in\mathcal{K}} \mathrm{dist}(f,V_n)$ algorithm performance
- Buffa, Maday, Patera, Prud'home, Turinici (BMPPT)
- BMPPT prove that for the ideal greedy $\sigma_n(\mathcal{K}) \le Cn2^n d_n(\mathcal{K})$
- Only useful if $d_n(\mathcal{K}) \to 0$ exponentially fast

Improved Convergence Results

- Binev, Cohen, Dahmen, DeVore, Petrova, Wojtaczszyk

Improved Convergence Results

- Binev, Cohen, Dahmen, DeVore, Petrova, Wojtaczszyk
- The $Cn2^n$ in BMPPT can be improved to 2^n and this cannot be improved

Improved Convergence Results

- Binev, Cohen, Dahmen, DeVore, Petrova, Wojtaczszyk
- The $Cn2^n$ in BMPPT can be improved to 2^n and this cannot be improved
- However, there are more optimistic results

Improved Convergence Results

- Binev, Cohen, Dahmen, DeVore, Petrova, Wojtaczszyk
- The $Cn2^n$ in BMPPT can be improved to 2^n and this cannot be improved
- However, there are more optimistic results
- These apply to both greedy and weak greedy

Improved Convergence Results

- Binev, Cohen, Dahmen, DeVore, Petrova, Wojtaczszyk
- The $Cn2^n$ in BMPPT can be improved to 2^n and this cannot be improved
- However, there are more optimistic results
- These apply to both greedy and weak greedy
- Slow Decay Theorem: If $d_n(\mathcal{K}) \le Mn^{-\alpha}, \quad n \ge 1$, then $\sigma_n(\mathcal{K}) \le C_\alpha M n^{-\alpha}, \quad n \ge 1$, with a fixed C_α

Improved Convergence Results

- Binev, Cohen, Dahmen, DeVore, Petrova, Wojtaczszyk
- The $Cn2^n$ in BMPPT can be improved to 2^n and this cannot be improved
- However, there are more optimistic results
- These apply to both greedy and weak greedy
- Slow Decay Theorem: If $d_n(\mathcal{K}) \leq Mn^{-\alpha}, \quad n \geq 1$, then $\sigma_n(\mathcal{K}) \leq C_\alpha Mn^{-\alpha}, \quad n \geq 1$, with a fixed C_α
- So for slow decay the (weak) greedy algorithm is near optimal.

Sub exponential rates

- BCDDPW also prove comparisons for subexponential decay

Sub exponential rates

- BCDDPW also prove comparisons for subexponential decay
- Subexponential Decay Theorem: If

$$d_n(\mathcal{K}) \leq Me^{-cn^\alpha}, \quad n = 1, 2, \dots,$$

then

$$\sigma_n(\mathcal{K}) \leq CMe^{-c'n^\beta} \quad n = 1, 2, \dots,$$

with a fixed C, c' and $\beta := \frac{\alpha}{\alpha+1}$

Sub exponential rates

- BCDDPW also prove comparisons for subexponential decay
- Subexponential Decay Theorem: If

$$d_n(\mathcal{K}) \le Me^{-cn^\alpha}, \quad n = 1, 2, \dots,$$

then

$$\sigma_n(\mathcal{K}) \le CMe^{-c'n^\beta} \quad n = 1, 2, \dots,$$

with a fixed C, c' and $\beta := \frac{\alpha}{\alpha+1}$

- For $\alpha < 1$, this result is stronger than BMPPT

Sub exponential rates

- BCDDPW also prove comparisons for subexponential decay
- Subexponential Decay Theorem: If

$$d_n(\mathcal{K}) \leq M e^{-cn^\alpha}, \quad n = 1, 2, \ldots,$$

then

$$\sigma_n(\mathcal{K}) \leq CM e^{-c'n^\beta} \quad n = 1, 2, \ldots,$$

with a fixed C, c' and $\beta := \frac{\alpha}{\alpha+1}$

- For $\alpha < 1$, this result is stronger than BMPPT
- For $\alpha > 1$, BMPPT is stronger

Sub exponential rates

- BCDDPW also prove comparisons for subexponential decay
- Subexponential Decay Theorem: If

$$d_n(\mathcal{K}) \leq M e^{-cn^\alpha}, \quad n = 1, 2, \dots,$$

then

$$\sigma_n(\mathcal{K}) \leq C M e^{-c' n^\beta} \quad n = 1, 2, \dots,$$

with a fixed C, c' and $\beta := \frac{\alpha}{\alpha+1}$

- For $\alpha < 1$, this result is stronger than BMPPT
- For $\alpha > 1$, BMPPT is stronger
- For $\alpha = 1$, depends on c

Robustness

- In Numerical Implementation the f_j are not computed exactly - space discretization in $\mathcal{H}_0^1$

Robustness

- In Numerical Implementation the f_j are not computed exactly - space discretization in $\mathcal{H}_0^1$
- So we consider a noisy version of WGA

Robustness

- In Numerical Implementation the f_j are not computed exactly - space discretization in $\mathcal{H}_0^1$
- So we consider a noisy version of WGA
- $f_0 \in \mathcal{K}$ chosen so that $\|f_0\| \geq \gamma \sup_{f \in \mathcal{K}} \|f\|$

Robustness

- In Numerical Implementation the f_j are not computed exactly - space discretization in $\mathcal{H}_0^1$
- So we consider a noisy version of WGA
- $f_0 \in \mathcal{K}$ chosen so that $\|f_0\| \geq \gamma \sup_{f \in \mathcal{K}} \|f\|$
- Approximated by $\hat{f}_0$ (generally not in K): $\|f_0 - \hat{f}_0\| \leq \epsilon$

Robustness

- In Numerical Implementation the f_j are not computed exactly - space discretization in $\mathcal{H}_0^1$
- So we consider a noisy version of WGA
- $f_0 \in \mathcal{K}$ chosen so that $\|f_0\| \geq \gamma \sup_{f \in \mathcal{K}} \|f\|$
- Approximated by $\hat{f}_0$ (generally not in K): $\|f_0 - \hat{f}_0\| \leq \epsilon$
- $\hat{F}_n := \mathrm{span}\{\hat{f}_0, \ldots, \hat{f}_{n-1}\}$

Robustness

- In Numerical Implementation the f_j are not computed exactly - space discretization in $\mathcal{H}_0^1$
- So we consider a noisy version of WGA
- $f_0 \in \mathcal{K}$ chosen so that $\|f_0\| \geq \gamma \sup_{f \in \mathcal{K}} \|f\|$
- Approximated by $\hat{f}_0$ (generally not in K): $\|f_0 - \hat{f}_0\| \leq \epsilon$
- $\hat{F}_n := \text{span}\{\hat{f}_0, \ldots, \hat{f}_{n-1}\}$
- $f_n \in \mathcal{K}$ chosen so that $\text{dist}(f_n, \hat{F}_n) \geq \gamma \sup_{f \in \mathcal{K}} \text{dist}(f, \hat{F}_n)$

Robustness

- In Numerical Implementation the f_j are not computed exactly - space discretization in $\mathcal{H}_0^1$
- So we consider a noisy version of WGA
- $f_0 \in \mathcal{K}$ chosen so that $\|f_0\| \geq \gamma \sup_{f \in \mathcal{K}} \|f\|$
- Approximated by $\hat{f}_0$ (generally not in K): $\|f_0 - \hat{f}_0\| \leq \epsilon$
- $\hat{F}_n := \mathrm{span}\{\hat{f}_0, \ldots, \hat{f}_{n-1}\}$
- $f_n \in \mathcal{K}$ chosen so that $\mathrm{dist}(f_n, \hat{F}_n) \geq \gamma \sup_{f \in \mathcal{K}} \mathrm{dist}(f, \hat{F}_n)$
- Approximate by $\hat{f}_n$ with $\|f_n - \hat{f}_n\| \leq \epsilon$

Performance of Noisy WGA

- Define

$$\hat{\sigma}_n(\mathcal{K}) := \sup_{f \in \mathcal{K}} \operatorname{dist}(f, \hat{F}_n)$$

Performance of Noisy WGA

- Define

$$\hat{\sigma}_n(\mathcal{K}) := \sup_{f\in\mathcal{K}} \mathrm{dist}(f, \hat{F}_n)$$

- THEOREM: If for some $\alpha > 0$

$$d_n(\mathcal{K}) \le M n^{-\alpha}, \quad n = 1, 2, \ldots$$

then

$$\sigma_n(\mathcal{K}) \le C(\alpha, \gamma) \max\{M n^{-\alpha}, \epsilon\}, \quad n = 1, 2, \ldots .$$

Performance of Noisy WGA

- Define

$$\hat{\sigma}_n(\mathcal{K}) := \sup_{f\in\mathcal{K}} \operatorname{dist}(f, \hat{F}_n)$$

- THEOREM: If for some $\alpha > 0$

$$d_n(\mathcal{K}) \leq Mn^{-\alpha}, \quad n = 1, 2, \dots$$

then

$$\sigma_n(\mathcal{K}) \leq C(\alpha,\gamma)\max\{Mn^{-\alpha}, \epsilon\}, \quad n = 1, 2, \dots.$$

- Similar Result holds for sub-exponential rates

Matrix Formulation

- We shall continue our discussion only for the strong greedy algorithm - WOLOG $H = \ell_2(\mathbb{N})$

Matrix Formulation

- We shall continue our discussion only for the strong greedy algorithm - WOLOG $H = \ell_2(\mathbb{N})$
- The greedy algorithm when applied to any set $\mathcal{K}$ leads to a canonical lower triangular matrix $A = A_{\mathcal{K}}$

Matrix Formulation

- We shall continue our discussion only for the strong greedy algorithm - WOLOG $H = \ell_2(\mathbb{N})$
- The greedy algorithm when applied to any set $\mathcal{K}$ leads to a canonical lower triangular matrix $A = A_{\mathcal{K}}$
- If $f_0, f_1, \dots,$ are the greedy selection and $f_0^*, f_1^*, \dots$ are their Gram-Schmidt orthogonalization then $f_i = \sum_{j=0}^{i} a_{ij} f_j^*$

$$A = \begin{pmatrix} a_{00} & 0 & 0 & \cdots & 0 & 0 & 0 & \cdots \\ a_{10} & a_{11} & 0 & \cdots & 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n0} & a_{n1} & a_{n2} & \cdots & a_{nn} & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}$$

Properties of A

- $A = (a_{ij})$ has many remarkable properties

$$\left(\begin{array}{ccc|ccccc} a_{\ell 0} & \cdots & a_{\ell\ell-1} & a_{\ell\ell} & 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{i0} & \cdots & a_{i\ell-1} & a_{i\ell} & \cdots & a_{ii} & 0 & \cdots \end{array}\right)$$

$$V_\ell$$

Properties of A

- $A = (a_{ij})$ has many remarkable properties
  - $|a_{nn}| = \sigma_n$

$$
\left(\begin{array}{ccc|ccccc}
a_{\ell 0} & \cdots & a_{\ell\ell-1} & a_{\ell\ell} & 0 & 0 & 0 & \cdots \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\
a_{i0} & \cdots & a_{i\ell-1} & a_{i\ell} & \cdots & a_{ii} & 0 & \cdots
\end{array}\right)
$$

$$V_\ell$$

Properties of A

- $A = (a_{ij})$ has many remarkable properties
  - $|a_{nn}| = \sigma_n$
  - For each $\ell = 1, 2, \ldots$

$$\sum_{j=\ell}^{i} a_{ij}^2 \leq a_{\ell\ell}^2 = \sigma_\ell^2, \quad i \geq \ell$$

$$\left(\begin{array}{ccc|ccccc} a_{\ell 0} & \cdots & a_{\ell\ell-1} & a_{\ell\ell} & 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{i0} & \cdots & a_{i\ell-1} & a_{i\ell} & \cdots & a_{ii} & 0 & \cdots \end{array}\right)$$

$$V_\ell$$

Properties of A

- $A = (a_{ij})$ has many remarkable properties
  - $|a_{nn}| = \sigma_n$
  - For each $\ell = 1, 2, \ldots$

$$\sum_{j=\ell}^{i} a_{ij}^2 \le a_{\ell\ell}^2 = \sigma_\ell^2, \quad i \ge \ell$$

$$\left(\begin{array}{ccc|ccccc} a_{\ell 0} & \cdots & a_{\ell\ell-1} & a_{\ell\ell} & 0 & 0 & 0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{i0} & \cdots & a_{i\ell-1} & a_{i\ell} & \cdots & a_{ii} & 0 & \cdots \end{array}\right)$$

$$V_\ell$$

- The Greedy Algorithm for $\mathcal{K}$ is equivalent to greedy for the rows of A: we can assume $\mathcal{K}$ is the set of rows of A and the $f_j^* = e_j$ the coordinate basis

Proof of Slow Decay Theorem

- The proofs of the main results are based on the following

Proof of Slow Decay Theorem

- The proofs of the main results are based on the following
- Flatness Lemma: If greedy error does not decay very fast then the Kolmogorov width must be large

Proof of Slow Decay Theorem

- The proofs of the main results are based on the following
- Flatness Lemma: If greedy error does not decay very fast then the Kolmogorov width must be large
- Suppose $\sigma_{n+2m} \geq \theta\sigma_n$ for $\theta = \frac{2\sqrt{2}}{3}$ and some n, m

Proof of Slow Decay Theorem

- The proofs of the main results are based on the following
- Flatness Lemma: If greedy error does not decay very fast then the Kolmogorov width must be large
- Suppose $\sigma_{n+2m} \geq \theta\sigma_n$ for $\theta = \frac{2\sqrt{2}}{3}$ and some n, m
- Then $\sigma_n \leq 3d_m$

Proof of Slow Decay Theorem

- The proofs of the main results are based on the following
- Flatness Lemma: If greedy error does not decay very fast then the Kolmogorov width must be large
- Suppose $\sigma_{n+2m} \geq \theta\sigma_n$ for $\theta = \frac{2\sqrt{2}}{3}$ and some n, m
- Then $\sigma_n \leq 3d_m$
- Notice that sequences like n^{-s} satisfy this flatness with $m \geq cn$

Proof of Slow Decay Theorem

- The proofs of the main results are based on the following
- Flatness Lemma: If greedy error does not decay very fast then the Kolmogorov width must be large
- Suppose $\sigma_{n+2m} \geq \theta\sigma_n$ for $\theta = \frac{2\sqrt{2}}{3}$ and some n, m
- Then $\sigma_n \leq 3d_m$
- Notice that sequences like n^{-s} satisfy this flatness with $m \geq cn$
- A similar result holds for any other choice of θ provided the constant 3 is changed suitably

Proof of Slow Decay Theorem

- We know $d_n \leq Mn^{-\alpha}, \ n = 1, 2, \ldots$ and want to prove $\sigma_n \leq C_\alpha M n^{-\alpha}$ for some C_α

Proof of Slow Decay Theorem

- We know $d_n \leq Mn^{-\alpha}, \; n = 1, 2, \ldots$ and want to prove $\sigma_n \leq C_\alpha M n^{-\alpha}$ for some C_α
- Define $\mu := 1 - \theta^{1/\alpha} > 0$ and $C_\alpha := 3(\mu/2)^{-\alpha}$

Proof of Slow Decay Theorem

- We know $d_n \leq Mn^{-\alpha}, \ n = 1, 2, \ldots$ and want to prove $\sigma_n \leq C_\alpha M n^{-\alpha}$ for some C_α
- Define $\mu := 1 - \theta^{1/\alpha} > 0$ and $C_\alpha := 3(\mu/2)^{-\alpha}$
- We suppose there is a N such that $\sigma_N > C_\alpha M N^{-\alpha}$ and go for a contradiction: Choose N smallest

Proof of Slow Decay Theorem

- We know $d_n \le M n^{-\alpha}, \; n = 1, 2, \ldots$ and want to prove $\sigma_n \le C_\alpha M n^{-\alpha}$ for some C_α
- Define $\mu := 1 - \theta^{1/\alpha} > 0$ and $C_\alpha := 3(\mu/2)^{-\alpha}$
- We suppose there is a N such that $\sigma_N > C_\alpha M N^{-\alpha}$ and go for a contradiction: Choose N smallest
- Then
$\sigma_n \le C_\alpha M n^{-\alpha} \le C_\alpha M (\frac{N}{n})^\alpha N^{-\alpha} \le (\frac{N}{n})^\alpha \sigma_N, \;\; n \le N$

Proof of Slow Decay Theorem

- We know $d_n \leq Mn^{-\alpha}, \ n = 1, 2, \ldots$ and want to prove $\sigma_n \leq C_\alpha M n^{-\alpha}$ for some C_α
- Define $\mu := 1 - \theta^{1/\alpha} > 0$ and $C_\alpha := 3(\mu/2)^{-\alpha}$
- We suppose there is a N such that $\sigma_N > C_\alpha M N^{-\alpha}$ and go for a contradiction: Choose N smallest
- Then $\sigma_n \leq C_\alpha M n^{-\alpha} \leq C_\alpha M (\frac{N}{n})^\alpha N^{-\alpha} \leq (\frac{N}{n})^\alpha \sigma_N, \ \ n \leq N$
- For $n = N - \mu N$ we have $\frac{\sigma_N}{\sigma_n} \geq (\frac{n}{N})^\alpha = (1 - \mu)^\alpha = \theta$

Proof of Slow Decay Theorem

- We know $d_n \leq Mn^{-\alpha}, \; n = 1, 2, \ldots$ and want to prove $\sigma_n \leq C_\alpha M n^{-\alpha}$ for some C_α
- Define $\mu := 1 - \theta^{1/\alpha} > 0$ and $C_\alpha := 3(\mu/2)^{-\alpha}$
- We suppose there is a N such that $\sigma_N > C_\alpha M N^{-\alpha}$ and go for a contradiction: Choose N smallest
- Then $\sigma_n \leq C_\alpha M n^{-\alpha} \leq C_\alpha M (\frac{N}{n})^\alpha N^{-\alpha} \leq (\frac{N}{n})^\alpha \sigma_N, \;\; n \leq N$
- For $n = N - \mu N$ we have $\frac{\sigma_N}{\sigma_n} \geq (\frac{n}{N})^\alpha = (1-\mu)^\alpha = \theta$
- We found a flat stretch for $\sigma_n(\mathcal{K})$

Proof of Slow Decay Theorem

- We know $d_n \leq Mn^{-\alpha}, \;\; n = 1, 2, \ldots$ and want to prove $\sigma_n \leq C_\alpha M n^{-\alpha}$ for some C_α
- Define $\mu := 1 - \theta^{1/\alpha} > 0$ and $C_\alpha := 3(\mu/2)^{-\alpha}$
- We suppose there is a N such that $\sigma_N > C_\alpha M N^{-\alpha}$ and go for a contradiction: Choose N smallest
- Then $\sigma_n \leq C_\alpha M n^{-\alpha} \leq C_\alpha M (\frac{N}{n})^\alpha N^{-\alpha} \leq (\frac{N}{n})^\alpha \sigma_N, \;\; n \leq N$
- For $n = N - \mu N$ we have $\frac{\sigma_N}{\sigma_n} \geq (\frac{n}{N})^\alpha = (1-\mu)^\alpha = \theta$
- We found a flat stretch for $\sigma_n(\mathcal{K})$
- So flatness lemma with $m = \mu N/2$ gives $\sigma_N \leq \sigma_n \leq 3 d_{\mu N} \leq 3M(\mu N)^{-\alpha} \leq C_\alpha M N^{-\alpha}$

Proof of Slow Decay Theorem

- We know $d_n \le Mn^{-\alpha}, \;\; n = 1, 2, \ldots$ and want to prove $\sigma_n \le C_\alpha M n^{-\alpha}$ for some C_α
- Define $\mu := 1 - \theta^{1/\alpha} > 0$ and $C_\alpha := 3(\mu/2)^{-\alpha}$
- We suppose there is a N such that $\sigma_N > C_\alpha M N^{-\alpha}$ and go for a contradiction: Choose N smallest
- Then $\sigma_n \le C_\alpha M n^{-\alpha} \le C_\alpha M (\frac{N}{n})^\alpha N^{-\alpha} \le (\frac{N}{n})^\alpha \sigma_N, \;\; n \le N$
- For $n = N - \mu N$ we have $\frac{\sigma_N}{\sigma_n} \ge (\frac{n}{N})^\alpha = (1-\mu)^\alpha = \theta$
- We found a flat stretch for $\sigma_n(\mathcal{K})$
- So flatness lemma with $m = \mu N/2$ gives $\sigma_N \le \sigma_n \le 3 d_{\mu N} \le 3M(\mu N)^{-\alpha} \le C_\alpha M N^{-\alpha}$
- This is the desired contradiction

Proof of Flatness

- We assume $\sigma_{n+2m} \geq \theta\sigma_n$ and estimate d_m

Proof of Flatness

- We assume $\sigma_{n+2m} \geq \theta\sigma_n$ and estimate d_m
- A m dimensional space cannot approximate the $2m+1$ coordinate vectors $e_n, \dots, e_{n+2m}$ well

Proof of Flatness

- We assume $\sigma_{n+2m} \geq \theta\sigma_n$ and estimate d_m
- A m dimensional space cannot approximate the $2m+1$ coordinate vectors $e_n, \ldots, e_{n+2m}$ well
- Let V_m^* be best Kolmogorov space of dimension m for $\mathcal{K}$ and consider an orthonormal basis $\phi_1, \ldots, \phi_m$ of V_m^*

$$\begin{pmatrix} \langle e_{n+1}, \phi_1 \rangle^2 & \langle e_{n+1}, \phi_2 \rangle^2 & \cdots & \langle e_{n+1}, \phi_m \rangle^2 \\ \cdots & \cdots & \cdots & \cdots \\ \langle e_{n+2m}, \phi_1 \rangle^2 & \langle e_{n+2m}, \phi_2 \rangle^2 & \cdots & \langle e_{n+2m}, \phi_m \rangle^2 \end{pmatrix}$$

Proof of Flatness

- We assume $\sigma_{n+2m} \geq \theta\sigma_n$ and estimate d_m
- A m dimensional space cannot approximate the $2m+1$ coordinate vectors $e_n, \dots, e_{n+2m}$ well
- Let V_m^* be best Kolmogorov space of dimension m for $\mathcal{K}$ and consider an orthonormal basis $\phi_1, \dots, \phi_m$ of V_m^*

$$\begin{pmatrix} \langle e_{n+1}, \phi_1\rangle^2 & \langle e_{n+1}, \phi_2\rangle^2 & \cdots & \langle e_{n+1}, \phi_m\rangle^2 \\ \cdots & \cdots & \cdots & \cdots \\ \langle e_{n+2m}, \phi_1\rangle^2 & \langle e_{n+2m}, \phi_2\rangle^2 & \cdots & \langle e_{n+2m}, \phi_m\rangle^2 \end{pmatrix}$$

- Column sums ≤ 1 and hence one row sum is $\leq 1/2$

Proof of Flatness

- We assume $\sigma_{n+2m} \geq \theta\sigma_n$ and estimate d_m
- A m dimensional space cannot approximate the $2m+1$ coordinate vectors $e_n, \dots, e_{n+2m}$ well
- Let V_m^* be best Kolmogorov space of dimension m for $\mathcal{K}$ and consider an orthonormal basis $\phi_1, \dots, \phi_m$ of V_m^*

$$\begin{pmatrix} \langle e_{n+1}, \phi_1\rangle^2 & \langle e_{n+1}, \phi_2\rangle^2 & \cdots & \langle e_{n+1}, \phi_m\rangle^2 \\ \cdots & \cdots & \cdots & \cdots \\ \langle e_{n+2m}, \phi_1\rangle^2 & \langle e_{n+2m}, \phi_2\rangle^2 & \cdots & \langle e_{n+2m}, \phi_m\rangle^2 \end{pmatrix}$$

- Column sums ≤ 1 and hence one row sum is $\leq 1/2$
- For one value of $j \in \{n+1, \dots, n+2m\}$ we have

$$\mathrm{dist}(e_j, V_m^*) \geq 1/\sqrt{2}$$

Proof of Flatness Lemma continued

- For this value of j consider $\hat{f}_j = a_j + \sigma_j e_j$ the restriction of the j-th row of A to $\{n, \ldots, n+2m\}$

$$\begin{matrix} a_{n,n} & 0 & \cdots & \cdots & \cdots & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{j,n} & \cdots & a_{j,j-1} & a_{j,j} & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ a_{n+2m,n} & \cdots & \cdots & \cdots & \cdots & \cdots & a_{n+2m,n+2m} \end{matrix}$$

- By property of A we have $\|a_j\|^2 \leq \sigma_n^2 - \sigma_j^2 \leq (1-\theta^2)\sigma_n^2$
- $\theta\sigma_n \leq \sigma_j \leq \sqrt{2}\mathrm{dist}(\sigma_j e_j, V_m^*) \leq \sqrt{2}d_m + \sqrt{2}\sqrt{1-\theta^2}\sigma_n$
- Conclude by bringing σ_n from right side to left

Concluding Remarks

- Should we be greedy?

Concluding Remarks

- Should we be greedy?
  - Momma always told me do not be greedy

Concluding Remarks

- Should we be greedy?
  - Momma always told me do not be greedy
  - Yvon Maday says dont listen to your Momma

Concluding Remarks

- Should we be greedy?
  - Momma always told me do not be greedy
  - Yvon Maday says dont listen to your Momma
  - Momma said when a man tells you not to listen to your Momma - run as fast as you can

Concluding Remarks

- Should we be greedy?
  - Momma always told me do not be greedy
  - Yvon Maday says dont listen to your Momma
  - Momma said when a man tells you not to listen to your Momma - run as fast as you can
- Less seriously: The greedy algorithm performs remarkably well

Concluding Remarks

- Should we be greedy?
  - Momma always told me do not be greedy
  - Yvon Maday says dont listen to your Momma
  - Momma said when a man tells you not to listen to your Momma - run as fast as you can
- Less seriously: The greedy algorithm performs remarkably well
- However it may be that a more cautious strategy can do even better