recent developments of approximation theory and greedy algorithms

Peter Binev

Department of Mathematics
and
Interdisciplinary Mathematics Institute

University of South Carolina

Reduced Order Modeling in General Relativity

Pasadena, CA June 6-7, 2013

Outline

Polynomial Approximations

find the best L_2-approximation via polynomials to a function in an interval $[a,b]$

- space $X = L_2[a,b]$ of functions: $f \in X \leftrightsquigarrow \|f\| < \infty$
- norm $\|f\| = \|f\|_X = \|f\|_2 := \left(\int\limits_{[a,b]} |f(x)|^2 dx \right)^{\frac{1}{2}}$
- basis $\varphi_1 = 1, \ \varphi_2 = x, \ \varphi_3 = x^2, \ \ldots \ , \ \varphi_n = x^{n-1}, \ \ldots$
- space of polynomials of degree $n-1$ $\qquad \Phi_n := \mathrm{span}\{\varphi_1, \varphi_2, \ldots, \varphi_n\}$
- approximation $\quad p_n := \underset{p \in \Phi_n}{\mathrm{argmin}} \|f - p\|$
- representation $\quad p_n = \sum_{j=1}^{n} c_{n,j}(f) \varphi_j$
- in general, the coefficients $c_{n,j}(f)$ are not easy to find
- in this case we can use orthogonality $\rightsquigarrow$ Hilbert spaces

Polynomial Approximations

find the best L_2-approximation via polynomials to a function in a domain Ω

- space $X = L_2(\Omega)$ of functions: $f \in X \leftrightsquigarrow \|f\| < \infty$
- norm $\|f\| = \|f\|_X = \|f\|_2 := \left(\int\limits_\Omega |f(x)|^2 dx\right)^{\frac{1}{2}}$
- basis φ_1, φ_2, φ_3, ... , φ_n, ...
- space of polynomials of degree n $\Phi_n := \operatorname{span}\{\varphi_1, \varphi_2, ..., \varphi_n\}$
- approximation $p_n := \underset{p \in \Phi_n}{\operatorname{argmin}} \|f - p\|$
- representation $p_n = \sum_{j=1}^{n} c_{n,j}(f)\varphi_j$
- in general, the coefficients $c_{n,j}(f)$ are not easy to find
- in this case we can use orthogonality $\rightsquigarrow$ Hilbert spaces

Hilbert Space Setup

- Banach space X $\rightsquigarrow$ normed linear space $\rightsquigarrow$ $\|f\|_X$
- Hilbert space H $\rightsquigarrow$ Banach space with a scalar product $\rightsquigarrow$ $\langle f, g\rangle$
- $L_2(\Omega)$ is a Hilbert space $\rightsquigarrow$ $\langle f, g\rangle := \int_\Omega f(x)\overline{g(x)}\,dx$ $\quad \bar{g}$ - complex conjugation
- (induced) norm $\quad \|f\| := \Big(\langle f, f\rangle\Big)^{\frac{1}{2}}$
- orthogonality $\quad f \perp g \;\leftrightsquigarrow\; \langle f, g\rangle = 0$
- orthogonal basis $\quad \psi_n = \varphi_n + \sum_{j=1}^{n-1} q_j\varphi_j$ and $\psi_n \perp \Phi_{n-1}$ $\quad$ Gramm-Schmidt
- space of polynomials $\quad \Phi_n = \mathrm{span}\{\varphi_1, \varphi_2, ..., \varphi_n\} = \mathrm{span}\{\psi_1, \psi_2, ..., \psi_n\}$
- representation $\quad p_n = \sum_{j=1}^{n} C_j(f)\psi_j := \operatorname*{argmin}_{p\in\Phi_n} \|f - p\|$ $\quad$ C_j do not depend on n
- $C_j(f) := \dfrac{\langle f, \psi_j\rangle}{\langle \psi_j, \psi_j\rangle}$ $\rightsquigarrow$ in case $\|\psi_j\| = 1$, we have $p_n = \sum_{j=1}^{n} \langle f, \psi_j\rangle\psi_j$

Approximation in Hilbert Spaces

- orthonormal basis of H: $\psi_1, \ \psi_2, \ \psi_3, \ \dots \ , \ \psi_n, \ \dots \ \leftrightsquigarrow \ \{\psi_j\}_{j=1}^{\infty}$

  $\langle \psi_j, \psi_k \rangle = \delta_{j,k} := \begin{cases} 0 & \text{if } j \neq k \\ 1 & \text{if } j = k \end{cases} \quad \rightsquigarrow \quad H = \operatorname{span}\{\psi_j\}_{j=1}^{\infty}$

- linear approximation $$p_n = \sum_{j=1}^{n} \langle f, \psi_j \rangle \psi_j \in \Phi_n$$

- approximation error $$\|f - p_n\|^2 = \sum_{j=n+1}^{\infty} \left| \langle f, \psi_j \rangle \right|^2$$

  Parseval's Identity
  $$\|f\|^2 = \sum_{j=1}^{\infty} |\langle f, \psi_j \rangle|^2$$

- nonlinear approximation $$g_n = \sum_{j=1}^{n} \langle f, \psi_{k_j} \rangle \psi_{k_j}$$

- index set $\Lambda_n = \{k_1, k_2, \dots, k_n\} \subset \mathbb{N} \ \rightsquigarrow \ \Lambda_n = \Lambda_{n-1} \cup \{k_n\}$

- *How to find* Λ_n*?*

  $$\rightsquigarrow \quad \left|\langle f, \psi_{k_1} \rangle\right| \geq \left|\langle f, \psi_{k_2} \rangle\right| \geq \left|\langle f, \psi_{k_3} \rangle\right| \geq \dots$$

Nonlinear Approximation

given a basis $\{\psi_j\}_{j=1}^{\infty}$, choose any n elements from it and form the linear combinations $\sum_{j=1}^{n} C_j \psi_{k_j}$

- approximation class (not a space!) $\Sigma_n := \left\{ g = \sum_{k \in \Lambda} C_k \psi_k : \Lambda \subset \mathbb{N}, \#\Lambda \leq n \right\}$
- $\Sigma_n = \Sigma_n \left(\{\psi_j\}_{j=1}^{\infty} \right)$ can be defined for any basis in a Banach space X
- approximate $f \in X$ via functions from Σ_n
- best approximation $\sigma_n(f) := \inf_{g \in \Sigma_n} \|f - g\|$
- basic question: *how to find* $g_n \in \Sigma_n$ *such that* $\|f - g_n\| \leq C\sigma_n$ *?*

Note that although $\{\psi_j\}_{j=1}^{\infty}$ and $\{\varphi_j\}_{j=1}^{\infty}$ might yield the same polynomial spaces Φ_n, it is usually the case that $\Sigma_n \left(\{\psi_j\}_{j=1}^{\infty} \right) \neq \Sigma_n \left(\{\varphi_j\}_{j=1}^{\infty} \right)$ for all n and even the rates of σ_n can be completely different

Greedy Approximation

how to find efficiently $g_n \in \Sigma_n$ *that approximates* f *well?*

- the case of an orthonormal basis $\{\psi_j\}_{j=1}^{\infty}$ in a Hilbert space X

  incremental algorithm for finding $g_n = \sum_{k \in \Lambda_n} \langle f, \psi_k \rangle \psi_k$

  $\Lambda_0 = \emptyset$ and
  $\Lambda_j = \Lambda_{j-1} \cup \{k_j\}$, where $k_j = \underset{k \in \mathbb{N} \setminus \Lambda_{j-1}}{\operatorname{argmax}} \langle f, \psi_k \rangle = \underset{k \in \mathbb{N}}{\operatorname{argmax}} \langle f - g_{j-1}, \psi_k \rangle$

- for a general basis in a Banach space X, let $f \in X$ has the representation $f = \sum_{j=1}^{\infty} c_j(f) \psi_j$

  Greedy Algorithm : define $g_n = \sum_{k \in \Lambda_n} c_k(f) \psi_k$

  for $\Lambda_0 = \emptyset$ and $\Lambda_j = \Lambda_{j-1} \cup \{k_j\}$, where $k_j = \underset{k \in \mathbb{N} \setminus \Lambda_{j-1}}{\operatorname{argmax}} c_k(f)$

Note that in the general case g_n is no longer the best approximation from Σ_n to f

Greedy Basis

the bases, for which g_n *is a good approximation*

- Greedy Basis $\{\psi_j\}_{j=1}^{\infty}$ $\leftrightsquigarrow$ for any $f \in X$ the greedy approximation $g_n = g_n(f)$ to f satisfies $\|f - g_n\| \le G\sigma_n(f)$ with a constant G independent on f and n
- Unconditional Basis $\{\psi_j\}_{j=1}^{\infty}$ $\leftrightsquigarrow$ for any sign sequence $\{\theta_j\}_{j=1}^{\infty}$, $\theta_j = \pm 1$, the operator M_θ defined by $M_\theta\Big(\sum_{j=1}^{\infty} a_j\psi_j\Big) = \sum_{j=1}^{\infty} \theta_j a_j \psi_j$ is bounded
- Democratic Basis $\{\psi_j\}_{j=1}^{\infty}$ $\leftrightsquigarrow$ there exists a constant D such that for any two finite sets of indeces P and Q with the same cardinality $\#P = \#Q$ we have $\Big\|\sum_{k\in P}^{\infty} \psi_k\Big\| \le D\Big\|\sum_{k\in Q}^{\infty} \psi_k\Big\|$

Theorem [Konyagin, Temlyakov]
A basis is greedy if and only if it is unconditional and democratic.

Weak Greedy Algorithm

- often it is difficult (or even impossible) to find the maximizing element ψ_k
- settle for an element which is at least γ times the best with $0 < \gamma \leq 1$
- define $g_n(f) := \sum_{k \in \Lambda_n} c_k(f)\psi_k$

  for $\Lambda_0 = \emptyset$ and $\Lambda_j = \Lambda_{j-1} \cup \{k_j\}$, where $c_{k_j}(f) \geq \gamma \max_{k \in I\!N \setminus \Lambda_{j-1}} c_k(f)$

Theorem [Konyagin, Temlyakov]

For any greedy basis of a Banach space X and any $\gamma \in (0,1]$ there is a basis-specific constant $C(\gamma)$, independent on f and n, such that

$$\|f - g_n(f)\| \leq C(\gamma)\, \sigma_n(f)$$

General Greedy Strategy

- start with $g_0 = 0$ and $\Lambda_0 = \emptyset$
- set $j = 1$ and loop through the next items
- analyze the element $f - g_0$ with the possible improvements related to (some of) the elements from $\{\psi_k\}_{k=1}^{\infty}$ by calculating a decision functional $\lambda_j(f - g_{j-1}, \psi_k)$ for each possible ψ_k
  - in tree approximation the number of possible elements is bounded by (a multiple of) j
  - in the classical settings λ_j is usually related to $\inf_{k,C} \|f - g_{j-1} - C\psi_k\|$
- be greedy, use the element ψ_{k_j} with the largest λ_j or at least the one, for which $\lambda_j(f - g_{j-1}, \psi_{k_j}) \geq \gamma \sup_k \lambda_j(f - g_{j-1}, \psi_k)$
- set $\Lambda_j = \Lambda_{j-1} \cup \{k_j\}$
- calculate the next approximation g_j based on $\{\psi_k\}_{k \in \Lambda_j}$
  - in the classical settings g_j is found in the form $g_{j-1} + C\psi_{k_j}$
- set $j := j + 1$ and continue the loop

Greedy Algorithms for Dictionaries in Hilbert Spaces

$\{\psi_k\}_{k=1}^{\infty}$ is a dictionary (not a basis!) in a Hilbert space with $\|\psi_k\| = 1$

- **Pure Greedy Algorithm**

  $k_j := \operatorname*{argmax}_k |\langle f - g_{j-1}, \psi_k \rangle|$ and $g_j = g_{j-1} + \langle f - g_{j-1}, \psi_{k_j} \rangle \psi_{k_j}$

Greedy Algorithms for Dictionaries in Hilbert Spaces

$\{\psi_k\}_{k=1}^{\infty}$ is a dictionary (not a basis!) in a Hilbert space with $\|\psi_k\| = 1$

- **Pure Greedy Algorithm**

  $k_j := \underset{k}{\operatorname{argmax}} \; |\langle f - g_{j-1}, \psi_k \rangle|$ and $g_j = g_{j-1} + \langle f - g_{j-1}, \psi_{k_j} \rangle \psi_{k_j}$

- **Orthogonal Greedy Algorithm**

  $k_j := \underset{k}{\operatorname{argmax}} \; |\langle f - g_{j-1}, \psi_k \rangle|$ and $g_j = \mathcal{P}_{\{\psi_k\}_{k \in \Lambda_j}} f$

  where $\mathcal{P}_\Psi f$ is the orthogonal projection of f on the space $\operatorname{span}\{\Psi\}$

Greedy Algorithms for Dictionaries in Hilbert Spaces

$\{\psi_k\}_{k=1}^{\infty}$ is a dictionary (not a basis!) in a Hilbert space with $\|\psi_k\| = 1$

- **Pure Greedy Algorithm**
  $k_j := \operatorname*{argmax}_k |\langle f - g_{j-1}, \psi_k \rangle|$ and $g_j = g_{j-1} + \langle f - g_{j-1}, \psi_{k_j} \rangle \psi_{k_j}$
- **Orthogonal Greedy Algorithm**
  $k_j := \operatorname*{argmax}_k |\langle f - g_{j-1}, \psi_k \rangle|$ and $g_j = \mathcal{P}_{\{\psi_k\}_{k \in \Lambda_j}} f$
  where $\mathcal{P}_\Psi f$ is the orthogonal projection of f on the space $\mathrm{span}\{\Psi\}$
- **Weak Greedy Algorithm** with $0 < \gamma \leq 1$
  $|\langle f - g_{j-1}, \psi_{k_j} \rangle| \geq \gamma \sup_k |\langle f - g_{j-1}, \psi_k \rangle|$
  and $g_j = g_{j-1} + \langle f - g_{j-1}, \psi_{k_j} \rangle \psi_{k_j}$

Greedy Algorithms for Dictionaries in Hilbert Spaces

$\{\psi_k\}_{k=1}^{\infty}$ is a dictionary (not a basis!) in a Hilbert space with $\|\psi_k\| = 1$

- **Pure Greedy Algorithm**
  $k_j := \operatorname*{argmax}_k \, |\langle f - g_{j-1}, \psi_k \rangle|$ and $g_j = g_{j-1} + \langle f - g_{j-1}, \psi_{k_j} \rangle \psi_{k_j}$
- **Orthogonal Greedy Algorithm**
  $k_j := \operatorname*{argmax}_k \, |\langle f - g_{j-1}, \psi_k \rangle|$ and $g_j = \mathcal{P}_{\{\psi_k\}_{k \in \Lambda_j}} f$
  where $\mathcal{P}_\Psi f$ is the orthogonal projection of f on the space $\mathrm{span}\{\Psi\}$
- **Weak Greedy Algorithm** with $0 < \gamma \leq 1$
  $|\langle f - g_{j-1}, \psi_{k_j} \rangle| \geq \gamma \sup_k \, |\langle f - g_{j-1}, \psi_k \rangle|$
  and $g_j = g_{j-1} + \langle f - g_{j-1}, \psi_{k_j} \rangle \psi_{k_j}$
- **Weak Orthogonal Greedy Algorithm** with $0 < \gamma \leq 1$
  $|\langle f - g_{j-1}, \psi_{k_j} \rangle| \geq \gamma \sup_k \, |\langle f - g_{j-1}, \psi_k \rangle|$ and $g_j = \mathcal{P}_{\{\psi_k\}_{k \in \Lambda_j}} f$

Greedy Algorithms for Dictionaries in Hilbert Spaces

$\{\psi_k\}_{k=1}^{\infty}$ is a dictionary (not a basis!) in a Hilbert space with $\|\psi_k\| = 1$

- **Pure Greedy Algorithm**

  $k_j := \underset{k}{\operatorname{argmax}} \; |\langle f - g_{j-1}, \psi_k\rangle|$ and $g_j = g_{j-1} + \langle f - g_{j-1}, \psi_{k_j}\rangle \psi_{k_j}$

- **Orthogonal Greedy Algorithm**

  $k_j := \underset{k}{\operatorname{argmax}} \; |\langle f - g_{j-1}, \psi_k\rangle|$ and $g_j = \mathcal{P}_{\{\psi_k\}_{k\in\Lambda_j}} f$

  where $\mathcal{P}_\Psi f$ is the orthogonal projection of f on the space $\operatorname{span}\{\Psi\}$

- **Weak Greedy Algorithm** with $0 < \gamma \le 1$

  $|\langle f - g_{j-1}, \psi_{k_j}\rangle| \ge \gamma \sup_k \; |\langle f - g_{j-1}, \psi_k\rangle|$

  and $g_j = g_{j-1} + \langle f - g_{j-1}, \psi_{k_j}\rangle \psi_{k_j}$

- **Weak Orthogonal Greedy Algorithm** with $0 < \gamma \le 1$

  $|\langle f - g_{j-1}, \psi_{k_j}\rangle| \ge \gamma \sup_k \; |\langle f - g_{j-1}, \psi_k\rangle|$ and $g_j = \mathcal{P}_{\{\psi_k\}_{k\in\Lambda_j}} f$

more in [V. Temlyakov, Greedy Approximation, Cambridge University Press, 2011]

Two Additional Examles

- **Coarse-to-Fine Algorithms in Tree Approximation**
  - framework for adaptive partitioning strategies
  - g_n corresponds to a (binary) tree with complexity n
  - functionals λ_k are estimators of the local errors
  - search for k_j is limited to the leaves of the tree corresponding to g_{j-1}
  - greedy strategy does not work $\rightsquigarrow$ needs modifications
  - theoretical estimates ensure near-best approximation with essentially linear complexity

Two Additional Examles

- **Coarse-to-Fine Algorithms in Tree Approximation**
  - framework for adaptive partitioning strategies
  - g_n corresponds to a (binary) tree with complexity n
  - functionals λ_k are estimators of the local errors
  - search for k_j is limited to the leaves of the tree corresponding to g_{j-1}
  - greedy strategy does not work $\rightsquigarrow$ needs modifications
  - theoretical estimates ensure near-best approximation with essentially linear complexity
- **Greedy Approach to Reduced Basis Method**
  - the problem is to estimate a high dimensional parametric set via a low dimensional subspace
  - the error cannot be calculated efficiently and one has to settle with a calculation of a surrogate $\rightsquigarrow$ using weak greedy strategies is a must
  - the general comparison of the greedy approximation with the best approximation requires exponential constants $\rightsquigarrow$ should apply finer estimation techniques

Adaptive Approximation on Binary Partitions

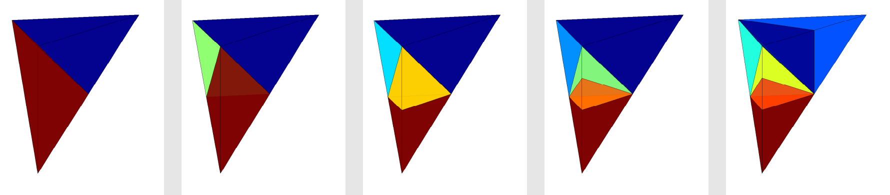
Function $f : \mathcal{X} \to \mathcal{Y}$

- $\mathcal{X} \subset I\!R^d$ domain equipped with a measure $\rho_{\mathcal{X}}$ such that $\rho_{\mathcal{X}}(\mathcal{X}) = 1$
- $\mathcal{Y} \subset [-M, M] \subset I\!R$ for a given constant M

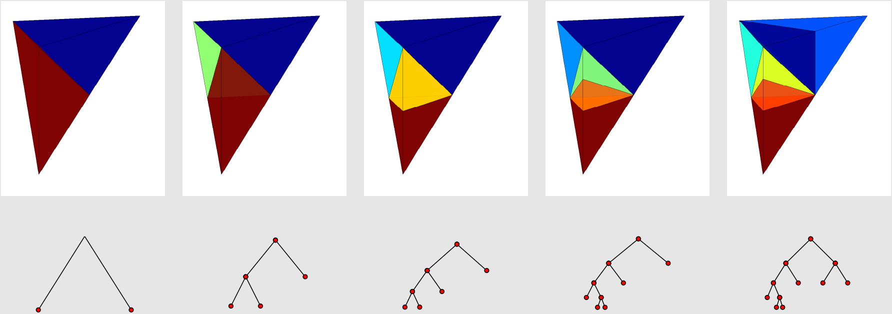
Adaptive binary partitions of $\mathcal{X}$

- building blocks $\Delta_{j,k}$ with $j = 1, 2, ...$ and $k = 0, 1, .., 2^j - 1$
- k represents a bitstream with length j
- $\Delta_{0,\emptyset} = \mathcal{X}$; $\Delta_{1,0} \cup \Delta_{1,1} = \Delta_{0,\emptyset}$ and $\rho_X(\Delta_{1,0} \cap \Delta_{1,1}) = 0$
- $\Delta_{j+1,2k} \cup \Delta_{j+1,2k+1} = \Delta_{j,k}$ and $\rho_{\mathcal{X}}(\Delta_{j+1,2k} \cap \Delta_{j+1,2k+1}) = 0$
- adaptive partition $\mathcal{P}$: *start with* $\Delta_{0,\emptyset}$ *and for certain pairs* (j,k) *replace* $\Delta_{j,k}$ *with* $\Delta_{j+1,2k}$ *and* $\Delta_{j+1,2k+1}$
- corresponding binary tree $\mathcal{T} = \mathcal{T}(\mathcal{P})$ with nodes $\Delta_{j,k}$

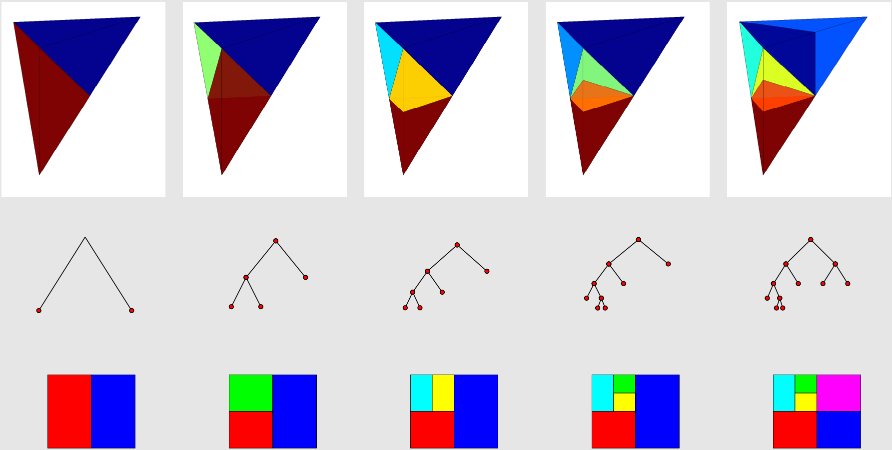
Binary Partitions



Binary Partitions



Binary Partitions



Adaptive Approximation on Binary Partitions

piecewise polynomial approximation of f on the partition $\mathcal{P}$

$$f_{\mathcal{P}}(x) := \sum_{\Delta \in \mathcal{P}} p_{\Delta,f}(x)\ \chi_{\Delta}(x)$$

- the process of finding an appropriate partition $\mathcal{P}$ can be defined on the corresponding tree $\mathcal{T} = \mathcal{T}(\mathcal{P})$ $\rightsquigarrow$ tree algorithms

Adaptive Approximation on Binary Partitions

piecewise polynomial approximation of f on the partition $\mathcal{P}$

$$f_{\mathcal{P}}(x) := \sum_{\Delta \in \mathcal{P}} p_{\Delta,f}(x)\ \chi_{\Delta}(x)$$

- the process of finding an appropriate partition $\mathcal{P}$ can be defined on the corresponding tree $\mathcal{T} = \mathcal{T}(\mathcal{P})$ $\leadsto$ tree algorithms
- in $\mathcal{T}$ the node $\Delta_{j,k}$ is the "parent" of its "children" $\Delta_{j+1,2k}$ and $\Delta_{j+1,2k+1}$
- not every tree corresponds to a partition $\leadsto$ admissible trees
  $\mathcal{T}$ *is admissible if for each node* $\Delta \in \mathcal{T}$ *its "sibling" is also in* $\mathcal{T}$
- the elements of $\mathcal{P}$ are the terminal nodes of $\mathcal{T}$, its "leaves" $\mathcal{L}(\mathcal{T})$
- usually the complexity of $\mathcal{P}$ is measured by the number of its elements $N = \#\mathcal{P}$
- the number of nodes of the binary tree $\mathcal{T}(\mathcal{P})$ is equivalent measure since $\#\mathcal{T}(\mathcal{P}) = 2N - 1$

Near-Best Approximation

Best Approximation

$$\sigma_N(f) := \inf_{\mathcal{P} : \#\mathcal{P} \leq N} \|f - f_{\mathcal{P}}\|$$

Near-Best Approximation

Best Approximation

$$\sigma_N(f) := \inf_{\mathcal{P}\,:\,\#\mathcal{P}\leq N} \|f - f_{\mathcal{P}}\|$$

Approximation class $\mathcal{A}^s(X)$: $f \in \mathcal{A}^s(X) \;\leftrightsquigarrow\; \sigma_N(f) = \mathcal{O}(N^{-s/d})$

shows the asymptotic behavior of the approximation

note the dependence on the dimension $\rightsquigarrow$ *curse of dimensionality*

Usually, the theoretical results are given in terms of how the algorithms perform for functions from an approximation class $\rightsquigarrow$ does not provide any assurance about the performance for an individual function. **Can we do better?**

Near-Best Approximation $\rightsquigarrow$ $\tilde{f}$

there exist constants $C_1 < \infty$ *and* $c_2 > 0$ *such that* $\|f - \tilde{f}\| \leq C_1\, \sigma_{c_2 N}(f)$

Sometimes referred as **instance optimality**

Error Functionals

- a functional $e:$ node $\Delta \in T \rightarrow$ error $e(\Delta) \geq 0$
- total error $E(T) := \sum_{\Delta \in \mathcal{L}(T)} e(\Delta)$.

Subadditivity

For any node $\Delta \in T$ if $\mathcal{C}(\Delta)$ is the set of its children, then

$$\sum_{\Delta' \in \mathcal{C}(\Delta)} e(\Delta') \leq e(\Delta)$$

Weak Subadditivity

There exists $C_0 \geq 1$ such that for any $\Delta \in T$ and for any finite subtree $T_\Delta \subset T$ with root node Δ

$$\sum_{\Delta' \in \mathcal{L}(T_\Delta)} e(\Delta') \leq C_0 \, e(\Delta)$$

Greedy Strategy for Tree Approximation

Example: Approximation in $L_2[0,1]$ of a function f defined as linear combination of scaled Haar functions: $f(x) := A\mathcal{H}_{\Delta_0} + B \sum_{\Delta\in\mathcal{I}} \mathcal{H}_\Delta$

- $\Delta_0 := [0, 2^{-M}]$, where M is huge constant
- $\mathcal{I}$ set of 2^{k-1} dyadic subintervals of $[\frac{1}{2}, 1]$ with length 2^{-k}
- $\|\mathcal{H}_\Delta\|_{L_2[0,1]} = 1$ and $A = B + \varepsilon$ with $\varepsilon > 0$ arbitrarily small $(B > 0)$
- $e([0, 2^{-m}]) = A^2$ for $m \leq M$ and $e(\Delta) = B^2$ for $\Delta \in \mathcal{I}$

The greedy algorithm will first subdivide $[1/2, 1]$ and its descendants until we obtain the set of intervals $\mathcal{I}$. From then on it will subdivide $[0, 2^{-m}]$ for $m \leq M$ (the ancestors of Δ_0). After $N := 2^k + M - 2$ subdivisions, the greedy algorithm will give the tree T with error $E(T) = \|f\|_{L_2}^2 = A^2 + 2^k B^2$.

If we would have subdivided $[1/2, 1]$ and its descendants to dyadic level $k+1$, we would have used just $n := 2^{k+1}$ subdivisions and gotten an error $\sigma_n(f) = A^2$.

$$\sigma_{2^{k+1}} = A^2 < 2^{-k}(A^2 + 2^k B^2) = E(T) \quad \text{with} \quad \#T = 2^k + M - 2 \gg 2^{k+1}$$

Modified Greedy Strategy for Tree Approximation

- the standard greedy strategy does not work for tree approximation
- need a modification that will change the decision functional
- design modified error functionals to appropriately penalize the depth of subdivision
- use the greedy strategy based on these modified error functionals
- use dynamic instead of static decision functionals
- extensions of the algorithms for high dimensions - sparse occupancy trees

Basic Idea of Tree Algorithm **[B., DeVore 2004]**

For all of the nodes of the initial tree T_0 we define $\tilde{e}(\Delta) = e(\Delta)$.

Then, for each child Δ_j, $j = 1, \ldots, m(\Delta)$ of Δ

$$\tilde{e}(\Delta_j) := q(\Delta) := \frac{\sum_{j=1}^{m(\Delta)} e(\Delta_j)}{e(\Delta) + \tilde{e}(\Delta)} \, \tilde{e}(\Delta).$$

Note that $\tilde{e}$ is constant on the children of Δ.

Define the penalty terms $p(\Delta_j) := \dfrac{e(\Delta_j)}{\tilde{e}(\Delta_j)}$

The main property of $\tilde{e}$:

$$\sum_{j=1}^{m(\Delta)} p(\Delta_j) = p(\Delta) + 1 \, .$$

Adaptive Algorithm on Binary Trees [2007]

Modified Error $\tilde{e}$ **:**

- initial partition $\rightsquigarrow$ subtree $T_0 \subset T$, $\Delta \in T_0$: $\tilde{e}(\Delta) := e(\Delta)$
- for each child Δ_j of Δ : $\tilde{e}(\Delta_j) := \left(\frac{1}{e(\Delta_j)} + \frac{1}{\tilde{e}(\Delta)} \right)^{-1}$

Adaptive Tree Algorithm
(creates a sequence of trees T_j, $j = 1, 2, \ldots$):

- start with T_0
- subdivide leaves $\Delta \in \mathcal{L}(T_{j-1})$ with largest $\tilde{e}(\Delta)$ to produce T_j

To eliminate sorting, we can consider all $\tilde{e}(\Delta)$ *with* $2^\ell \leq \tilde{e}(\Delta) < 2^{\ell+1}$ *as if they are equally large.*

Near-Best Approximation on Binary Trees

Best Approximation

$$\sigma_N(f) := \inf_{\mathcal{P}\,:\,\#\mathcal{P}\le N} \|f - f_{\mathcal{P}}\|$$

Assume that the error functionals $e(\Delta) \ge 0$ satisfy the subadditivity condition. Then the adaptive tree algorithm that produces a tree T_N corresponding to a partition $\mathcal{P}_N$ with $N \ge n$ elements satisfies

$$E(T_N) = \|f - \tilde{f}_N\| \le \left(\frac{N}{N-n+1}\right)\sigma_n(f)$$

This gives the constant $C_1 = \frac{N}{(1-c_2)N+1}$ *for any chosen* $0 < c_2 \le 1$

in the general estimate $\|f - \tilde{f}_N\| \le C_1\,\sigma_{c_2 N}(f)$.

Parameter Dependent PDEs

- input parameters $\mu \in D \subset \mathbb{R}^p$
- differential operator $A_\mu : \mathcal{H} \to \mathcal{H}'$
- functional $\ell \in \mathcal{H}'$
- solution u_μ of $A_\mu u_\mu = \ell(u_\mu)$
- quantity of interest $I(\mu) = \ell(u_\mu)$ $\quad I_\mu \to \mathrm{opt}_{\mu \in D}$
- example:

$$\|\cdot\|_{\mathcal{H}}^2 = \|\cdot\|^2 = a_{\bar{\mu}}(\cdot,\cdot)$$

$$\langle A_\mu, v\rangle := a_\mu(u,v) = \sum_{j=1}^{p} \theta_j(\mu_j) \int_{\Omega_j} \nabla u \cdot \nabla v \; dx$$

uniform ellipticity: $c_1 \|v\|^2 \le a_\mu(v,v) \le C_1 \|v\|^2 \qquad v \in \mathcal{H}, \;\; \mu \in D$

[Y. Maday, T. Patera, G. Turicini, ...]

Reduced Basis $\rightsquigarrow$ *exploit sparsity*

- "solution manifold"

compact set $\mathcal{F} := \{u_\mu = A_\mu^{-1}\ell \ : \ \mu \in D\} \subset \mathcal{H}$

- offline:

compute f_0, f_1,..., f_{n-1} such that for $F_n := \mathrm{span}\{f_0, ..., f_{n-1}\}$

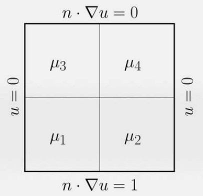
$$\sigma_n := \max_{f \in \mathcal{F}} \|f - P_n f\| \quad \leq \ [tollerance]$$

- online:

for each μ-query solve a small Galerkin problem in F_n

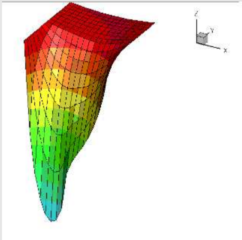
$$a_\mu(u_\mu^n, f_j) = \ell(f_j) \qquad\qquad j = 0, ..., n-1$$

Example



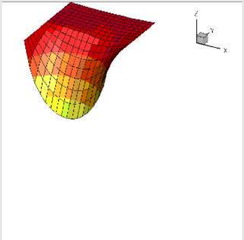
$$a_\mu(u,v) = \sum_{j=1}^{3} \int_{\Omega_j} \mu_j \, \nabla u \cdot \nabla v \; dx \; + \int_{\Omega_4} \nabla u \cdot \nabla v \; dx$$

$$\ell(v) = \int_{\Gamma_C} v \; ds$$



$\mu_1 = 0.1, \; \mu_2 = 0.3, \; \mu_3 = 0.8$

$\rightsquigarrow \quad \ell(u) = 1.24705$



$\mu_1 = 0.4, \; \mu_2 = 0.4, \; \mu_3 = 7.1$

$\rightsquigarrow \quad \ell(u) = 0.58505$

Reduced Basis $\rightsquigarrow$ *exploit sparsity*

- offline: f_0, f_1,..., f_{n-1} such that for $F_n := \mathrm{span}\{f_0, ..., f_{n-1}\}$

$$\sigma_n := \max_{f \in \mathcal{F}} \|f - P_n f\| \quad \leq \ [tollerance]$$

- online: for each μ solve a small Galerkin problem in F_n

$$a_\mu(u_\mu^n, f_j) = \ell(f_j) \qquad\qquad j = 0, ..., n-1$$

$\rightsquigarrow$

$$\begin{aligned} |\ell(u_\mu) - \ell(u_\mu^n)| &= a_\mu(u_\mu - u_\mu^n, u_\mu - u_\mu^n) \\ &\leq C_1\, \sigma_n(\mathcal{F})^2 \leq C_1\, [tollerance]^2 \end{aligned}$$

$\Downarrow$

- use u_μ^n for solving the optimization problem

Basis Construction $\rightsquigarrow$ *greedy approach*

- ideal algorithm $\rightsquigarrow$ pure greedy
  - $f_0 := \operatorname*{argmax}_{f \in \mathcal{F}} \|f\|$, $F_1 := \operatorname{span}\{f_0\}$, $\sigma_1(\mathcal{F}) := \|f_0\|$

  given $F_n := \operatorname{span}\{f_0, ..., f_{n-1}\}$ and $\sigma_n(\mathcal{F}) := \max_{f \in \mathcal{F}} \|f - P_n f\|$

  - $f_n := \operatorname*{argmax}_{f \in \mathcal{F}} \|f - P_n f\|$

- a feasible variant $\rightsquigarrow$ weak greedy algorithm

  using a computable "surrogate" $R_n(f)$ for which

$$c_2 R_n(f) \leq \|f - P_n f\| \leq C_2 R_n(f)$$

  - $\|f - f_n\| \geq \gamma \, \sigma_n(\mathcal{F})$
  - e.g. $f_n := \operatorname*{argmax}_{f \in \mathcal{F}} R_n(f)$ and $\gamma = \frac{c_2}{C_2}$

Kolmogorov Widths

$$d_n(\mathcal{F}) := \inf_{\dim(Y)=n} \sup_{f\in\mathcal{F}} \operatorname{dist}_{\mathcal{H}}(f, Y) \qquad \leq \sigma_n(\mathcal{F})$$

- Can one bound $\sigma_n(\mathcal{F})$ in terms of $d_n(\mathcal{F})$?

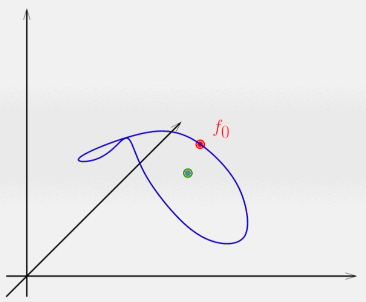
- Are optimal subspaces spanned by elements of $\mathcal{F}$?

$$\overline{d}_n(\mathcal{F}) := \inf_{Y\in\langle\mathcal{F}\rangle_n} \sup_{f\in\mathcal{F}} \operatorname{dist}_{\mathcal{H}}(f, Y) \qquad \leq \sigma_n(\mathcal{F})$$

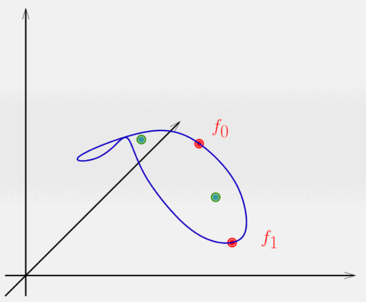
- Theorem:
  - for any compact set $\mathcal{F}$ we have $\overline{d}_n(\mathcal{F}) \leq (n+1)\, d_n(\mathcal{F})$
  - given any $\varepsilon > 0$ there is a set $\mathcal{F}$ such that

    $\overline{d}_n(\mathcal{F}) \geq (n-1-\varepsilon)\, d_n(\mathcal{F})$

Widths *vs* Greedy



$n = 1$



$n = 2$

Kolmogorov Widths *vs* Greedy Basis

Results $\rightsquigarrow$ *pure greedy* ($\gamma = 1$)

- [Buffa, Maday, Patera, Prudhomme, Turinici]

$$\sigma_n(\mathcal{F}) \leq C\ n2^n\ d_n(\mathcal{F})$$

- slight improvement

$$\sigma_n(\mathcal{F}) \leq \frac{2^{n+1}}{\sqrt{3}}\ d_n(\mathcal{F})$$

- for any $n > 0$ and any $\varepsilon > 0$ there exists a set $\mathcal{F} = \mathcal{F}_n$ such that for the pure greedy algorithm

$$\sigma_n(\mathcal{F}) \geq (1 - \varepsilon)2^n\ d_n(\mathcal{F})$$

Results $\rightsquigarrow$ *pure greedy* ($\gamma = 1$)

- [Buffa, Maday, Patera, Prudhomme, Turinici]

$$\sigma_n(\mathcal{F}) \le C \; n 2^n \; d_n(\mathcal{F})$$

- slight improvement

$$\sigma_n(\mathcal{F}) \le \frac{2^{n+1}}{\sqrt{3}} \; d_n(\mathcal{F})$$

- for any $n > 0$ and any $\varepsilon > 0$ there exists a set $\mathcal{F} = \mathcal{F}_n$ such that for the pure greedy algorithm

$$\sigma_n(\mathcal{F}) \ge (1 - \varepsilon) 2^n \; d_n(\mathcal{F})$$

- What if $2^n d_n(\mathcal{F}) \nrightarrow 0$?
- What if $\gamma < 1$?

Polynomial Convergence Rates

- **Theorem** [Binev, Cohen, Dahmen, DeVore, Petrova, Wojtaszczyk]

Suppose that $d_0(\mathcal{F}) \leq M$. *Then*

$d_n(\mathcal{F}) \leq Mn^{-\alpha}$ *for* $n > 0$ $\quad\Rightarrow\quad$ $\sigma_n(\mathcal{F}) \leq CMn^{-\alpha}$ *for* $n > 0$

with $C := 4^{\alpha} q^{\alpha + \frac{1}{2}}$ *and* $q := \lceil 2^{\alpha+1}\gamma^{-1} \rceil^2$.

Polynomial Convergence Rates

- **Theorem** [Binev, Cohen, Dahmen, DeVore, Petrova, Wojtaszczyk]

  Suppose that $d_0(\mathcal{F}) \leq M$. *Then*

  $d_n(\mathcal{F}) \leq M n^{-\alpha}$ *for* $n > 0$ $\quad\Rightarrow\quad$ $\sigma_n(\mathcal{F}) \leq C M n^{-\alpha}$ *for* $n > 0$

  with $C := 4^{\alpha} q^{\alpha+\frac{1}{2}}$ *and* $q := \lceil 2^{\alpha+1}\gamma^{-1}\rceil^2$.

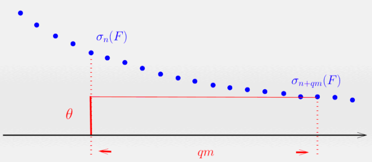
- using the "Flatness" Lemma:

  Let $0 < \theta < 1$ *and assume that for* $q := \lceil 2\theta^{-1}\gamma^{-1}\rceil^2$
  and some integers m *and* n *we have* $\sigma_{n+qm}(\mathcal{F}) \geq \theta\, \sigma_n(\mathcal{F})$.

  Then

$$\sigma_n(\mathcal{F}) \leq q^{\frac{1}{2}}\, d_m(\mathcal{F}).$$

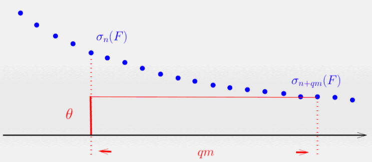
Idea of the Proof



$$\sigma_{n+qm}(\mathcal{F}) \geq \theta \sigma_n(\mathcal{F})$$

$$\Rightarrow \quad \sigma_n(\mathcal{F}) \leq q^{\frac{1}{2}} d_m(\mathcal{F})$$

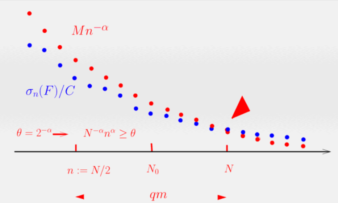
Idea of the Proof



$$\sigma_{n+qm}(\mathcal{F}) \geq \theta \sigma_n(\mathcal{F})$$

$$\Rightarrow \quad \sigma_n(\mathcal{F}) \leq q^{\frac{1}{2}} d_m(\mathcal{F})$$

- $\sigma_n(\mathcal{F}) \leq CMn^{-\alpha}$ for $n \leq N_0$
- assume it fails for some $N > N_0$
- $\Rightarrow$ flatness for $m \sim n$
- apply flatness lemma
- $\rightsquigarrow$ contradiction



$$d_n(\mathcal{F}) \leq Mn^{-\alpha} \text{ for } n > 0 \quad \Rightarrow \quad \sigma_n(\mathcal{F}) \leq CMn^{-\alpha} \text{ for } n > 0$$

Sub-Exponential Rates

- finer resolutions between $n^{-\alpha}$ and e^{-an}
- **Theorem** [DeVore, Petrova, Wojtaszczyk]

For any compact set $\mathcal{F}$ *and* $n \geq 1$, *we have*

$$\sigma_n(\mathcal{F}) \leq \frac{\sqrt{2}}{\gamma} \min_{1 \leq m < n} d_m^{\frac{n-m}{n}}(\mathcal{F})$$

In particular, $\sigma_{2n}(\mathcal{F}) \leq \frac{\sqrt{2 d_n(\mathcal{F})}}{\gamma}$ *and*

$$d_n(\mathcal{F}) \leq C_0 \, e^{-an^{\alpha}} \quad \textit{for} \quad n \geq 1$$

$$\Downarrow$$

$$\sigma_n(\mathcal{F}) \leq \frac{\sqrt{2C_0}}{\gamma} \, e^{-c_1 a n^{\alpha}} \quad \textit{for} \quad n \geq 1 \textit{ with } \; c_1 = 2^{-1-2\alpha} \, .$$

Robustness

- in reality f_j cannot be computed exactly $\rightsquigarrow$ we receive $\widetilde{f}_j$ *(that might not be in $\mathcal{F}$)* with $\|f_j - \widetilde{f}_j\| \leq \varepsilon$
- instead of F_n use $\widetilde{F}_n := \operatorname{span}\left\{\widetilde{f}_0, \dots, \widetilde{f}_{n-1}\right\}$
- performance of the noisy weak greedy algorithm

$$\widetilde{\sigma}_n(\mathcal{F}) := \sup_{f \in \mathcal{F}} \operatorname{dist}_{\mathcal{H}}(f, \widetilde{F}_n)$$

- Theorem [polynomial rates, $n > 0$]

$$d_n(\mathcal{F}) \leq M n^{-\alpha} \qquad \Rightarrow \qquad \widetilde{\sigma}_n(\mathcal{F}) \leq C \max\{M n^{-\alpha}, \varepsilon\}$$

with $C = C(\alpha, \gamma)$.

- similar result for subexponential rates

The End

THANK YOU!