Performance of greedy algorithm for Kolmogorov widths

P. Wojtaszczyk

University of Warsaw

Perspectives in Analysis and Probability
Conference in Honor of Freddy Delbaen
Zurich, September 24, 2012

Outline of the talk

Reduced basis

Let $F \subset \mathbb{R}^d$ be a compact set e.g. $F = [0,1]^d$ with $d = 50$ and let

$$D_\mu f = g \qquad \mu \in E \tag{1}$$

be a family of uniformly elliptic PDE's.
For a given $\mu \in F$ solving (1) is time consuming. We want to prepare ourselves to do it fast – on line. Idea of reduced basis method: (end of XX-century)
We solve (1) for $\mu_1, \ldots, \mu_n$ to get $f_{\mu_1}, \ldots, f_{\mu_n}$ and for given μ we approximate the solution

$$f_\mu \sim \sum_{j=1}^{n} a_j f_{\mu_j}.$$

How to find basis elements?

Let $F \subset \mathbb{R}^d$ be a compact set e.g. $F = [0,1]^d$ with $d = 50$ and let

$$D_\mu f = g \qquad \mu \in E \tag{1}$$

be a family of uniformly elliptic PDE's.
For a given $\mu \in F$ solving (1) is time consuming. We want to prepare ourselves to do it fast – on line. Idea of reduced basis method: (end of XX-century)
We solve (1) for $\mu_1, \ldots, \mu_n$ to get $f_{\mu_1}, \ldots, f_{\mu_n}$ and for given μ we approximate the solution

$$f_\mu \sim \sum_{j=1}^{n} a_j f_{\mu_j}.$$

How to find basis elements?

Reduced basis

Let $F \subset \mathbb{R}^d$ be a compact set e.g. $F = [0,1]^d$ with $d = 50$ and let

$$D_\mu f = g \qquad \mu \in E \tag{1}$$

be a family of uniformly elliptic PDE's.
For a given $\mu \in F$ solving (1) is time consuming. We want to prepare ourselves to do it fast – on line. Idea of reduced basis method: (end of XX-century)
We solve (1) for $\mu_1, \ldots, \mu_n$ to get $f_{\mu_1}, \ldots, f_{\mu_n}$ and for given μ we approximate the solution

$$f_\mu \sim \sum_{j=1}^{n} a_j f_{\mu_j}.$$

How to find basis elements?

Reduced basis

Let $F \subset \mathbb{R}^d$ be a compact set e.g. $F = [0,1]^d$ with $d = 50$ and let

$$D_\mu f = g \quad \mu \in E \tag{1}$$

be a family of uniformly elliptic PDE's.
For a given $\mu \in F$ solving (1) is time consuming. We want to prepare ourselves to do it fast – on line. Idea of reduced basis method: (end of XX-century)
We solve (1) for $\mu_1, \dots, \mu_n$ to get $f_{\mu_1}, \dots, f_{\mu_n}$ and for given μ we approximate the solution

$$f_\mu \sim \sum_{j=1}^{n} a_j f_{\mu_j}.$$

How to find basis elements?

Greedy selection of reduced basis, Maday-Patera-Turinici, 2002

This is a very theoretical version of the scheme.
Let $\mathcal{K} =: \{f_\mu \;:\; \mu \in F\}$ be a compact subset of certain Banach space $\mathcal{X}$ or a Hilbert space $\mathcal{H}$.
We define $\mu_1, \ldots \mu_n$ as follows ($f_j = f_{\mu_j}$)

1. $f_1 = \mathrm{argmax}\{\|f\| \;:\; f \in \mathcal{K}\}$
2. Given $f_1, \ldots, f_n$ we define $E_n = \mathrm{span}\,\{f_1, \ldots, f_n\}$ and put $f_{n+1} = \mathrm{argmax}\{\mathrm{dist}\,(f, E_n) \;:\; f \in \mathcal{K}\}$

We want to estimate the performance of this procedure.

We define $\sigma_n(\mathcal{K}) = \sup_{f\in\mathcal{K}} \mathrm{dist}\,(f, E_n)$.
Recall the Kolmogorov width

$$d_n(\mathcal{K}) = \inf_F \sup_{f\in\mathcal{K}} \mathrm{dist}(f, F)$$

where F is a subspace of dimension $\leq n$.

Clearly always $d_n(\mathcal{K}) \leq \sigma_n(\mathcal{K})$.

Greedy selection of reduced basis, Maday-Patera-Turinici, 2002

This is a very theoretical version of the scheme.
Let $\mathcal{K} =: \{f_\mu \ : \ \mu \in F\}$ be a compact subset of certain Banach space $\mathcal{X}$ or a Hilbert space $\mathcal{H}$.
We define $\mu_1, \dots \mu_n$ as follows ($f_j = f_{\mu_j}$)

1. $f_1 = \operatorname{argmax}\{\|f\| \ : \ f \in \mathcal{K}\}$
2. Given $f_1, \dots, f_n$ we define $E_n = \operatorname{span}\{f_1, \dots, f_n\}$ and put $f_{n+1} = \operatorname{argmax}\{\operatorname{dist}(f, E_n) \ : \ f \in \mathcal{K}\}$

We want to estimate the performance of this procedure.

We define $\sigma_n(\mathcal{K}) = \sup_{f\in\mathcal{K}} \operatorname{dist}(f, E_n)$.
Recall the Kolmogorov width

$$d_n(\mathcal{K}) = \inf_F \sup_{f\in\mathcal{K}} \operatorname{dist}(f, F)$$

where F is a subspace of dimension $\leq n$.

Clearly always $d_n(\mathcal{K}) \leq \sigma_n(\mathcal{K})$.

Greedy selection of reduced basis, Maday-Patera-Turinici, 2002

This is a very theoretical version of the scheme.
Let $\mathcal{K} =: \{f_\mu \ : \ \mu \in F\}$ be a compact subset of certain Banach space $\mathcal{X}$ or a Hilbert space $\mathcal{H}$.
We define $\mu_1, \dots \mu_n$ as follows ($f_j = f_{\mu_j}$)

1. $f_1 = \operatorname{argmax}\{\|f\| \ : \ f \in \mathcal{K}\}$
2. Given $f_1, \dots, f_n$ we define $E_n = \operatorname{span}\{f_1, \dots, f_n\}$ and put $f_{n+1} = \operatorname{argmax}\{\operatorname{dist}(f, E_n) \ : \ f \in \mathcal{K}\}$

We want to estimate the performance of this procedure.

We define $\sigma_n(\mathcal{K}) = \sup_{f \in \mathcal{K}} \operatorname{dist}(f, E_n)$.
Recall the Kolmogorov width

$$d_n(\mathcal{K}) = \inf_F \sup_{f \in \mathcal{K}} \operatorname{dist}(f, F)$$

where F is a subspace of dimension $\leq n$.

Clearly always $d_n(\mathcal{K}) \leq \sigma_n(\mathcal{K})$.

Greedy selection of reduced basis, Maday-Patera-Turinici, 2002

This is a very theoretical version of the scheme.
Let $\mathcal{K} =: \{f_\mu \ : \ \mu \in F\}$ be a compact subset of certain Banach space $\mathcal{X}$ or a Hilbert space $\mathcal{H}$.
We define $\mu_1, \dots \mu_n$ as follows ($f_j = f_{\mu_j}$)

1. $f_1 = \mathrm{argmax}\{\|f\| \ : \ f \in \mathcal{K}\}$
2. Given $f_1, \dots, f_n$ we define $E_n = \mathrm{span}\,\{f_1, \dots, f_n\}$ and put $f_{n+1} = \mathrm{argmax}\{\mathrm{dist}\,(f, E_n) \ : \ f \in \mathcal{K}\}$

We want to estimate the performance of this procedure.

We define $\sigma_n(\mathcal{K}) = \sup_{f\in\mathcal{K}} \mathrm{dist}\,(f, E_n)$.
Recall the Kolmogorov width

$$d_n(\mathcal{K}) = \inf_F \sup_{f\in\mathcal{K}} \mathrm{dist}(f, F)$$

where F is a subspace of dimension $\leq n$.

Clearly always $d_n(\mathcal{K}) \leq \sigma_n(\mathcal{K})$.

This talk reports

Results from two papers

BCDDPW Peter Binev, Albert Cohen, Wolfgang Dahmen, Ronald DeVore, Guergana Petrova, P.W., *Convergence Rates for Greedy Algorithms in Reduced Basis Methods* SIAM J. Math. Anal. vol.43 N.3 pp. 1457-1472 (2011)

DPW Ronald DeVore, Guergana Petrova, P.W., *Greedy Algorithms for Reduced Basis in Banach Spaces* submitted

Let us digress

Note that $f_n \in \mathcal{K}$. We define

$$\bar{d}_n(\mathcal{K}) = \inf_F \sup_{f \in \mathcal{K}} \mathrm{dist}(f, F)$$

where F is a subspace of dimension $\leq n$ spanned by elements from $\mathcal{K}$.

Theorem (BCDDPW)

The following holds:

(i) *For any compact set $\mathcal{F} \subset \mathcal{H}$ and any $n \geq 0$, we have $\bar{d}_n(\mathcal{F}) \leq (n+1) d_n(\mathcal{F})$.*

(ii) *Given any $n > 0$ and $\epsilon > 0$, there is a set $\mathcal{F} \subset \mathcal{H}$ such that $\bar{d}_n(\mathcal{F}) \geq (n-1-\epsilon) d_n(\mathcal{F})$.*

Let us digress

Note that $f_n \in \mathcal{K}$. We define

$$\bar{d}_n(\mathcal{K}) = \inf_F \sup_{f \in \mathcal{K}} \mathrm{dist}(f, F)$$

where F is a subspace of dimension $\leq n$ spanned by elements from $\mathcal{K}$.

Theorem (BCDDPW)

The following holds:

(i) *For any compact set $\mathcal{F} \subset \mathcal{H}$ and any $n \geq 0$, we have $\bar{d}_n(\mathcal{F}) \leq (n+1) d_n(\mathcal{F})$.*

(ii) *Given any $n > 0$ and $\epsilon > 0$, there is a set $\mathcal{F} \subset \mathcal{H}$ such that $\bar{d}_n(\mathcal{F}) \geq (n-1-\epsilon) d_n(\mathcal{F})$.*

Direct comparison

A.Buffa, Y.Maday, A.T.Patera, C.Prud'homme, G.Turinici, 2012, have shown that for $\mathcal{K} \subset \mathcal{H}$

$$\sigma_n(\mathcal{K}) \leq Cn2^n d_n(\mathcal{K}).$$

We have

Theorem (BCDDPW)

Let $\mathcal{F}$ be an arbitrary compact set in a Hilbert space $\mathcal{H}$. For each $n = 1, 2, \dots$ we have

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

For any $n > 0$, $\epsilon > 0$ there exists a set $\mathcal{K} = \mathcal{K}_{n,\epsilon}$ such that

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

This looks rather bad.

Direct comparison

A.Buffa, Y.Maday, A.T.Patera, C.Prud'homme, G.Turinici, 2012, have shown that for $\mathcal{K} \subset \mathcal{H}$

$$\sigma_n(\mathcal{K}) \leq Cn2^n d_n(\mathcal{K}).$$

We have

Theorem (BCDDPW)

Let $\mathcal{F}$ be an arbitrary compact set in a Hilbert space $\mathcal{H}$. For each $n = 1, 2, \ldots$ we have

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

For any $n > 0$, $\epsilon > 0$ there exists a set $\mathcal{K} = \mathcal{K}_{n,\epsilon}$ such that

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

This looks rather bad.

Matrix representation in $\mathcal{H}$

Algorithm produces $f_1, f_2, \ldots$. Let f_n^*, $n = 1, 2, \ldots$, be its Gram-Schmidt orthogonalization, so $P_n f = \sum_{i=1}^n \langle f, f_i^* \rangle f_i^*$, and in particular

$$f_i = P_i f_i = \sum_{j=1}^{i} a_{i,j} f_j^*, \quad a_{i,j} = \langle f_i, f_j^* \rangle,\ j \leq i. \tag{3}$$

The lower triangular matrix $A := A(\mathcal{F}) := (a_{i,j})_{i,j=1}^\infty, \quad a_{i,j} = 0, j > i$. contains all the information about the output of algorithm on $\mathcal{F}$. It satisfies

P1: The diagonal elements of A satisfy $|a_{n,n}| = \sigma_n := \sigma_n(\mathcal{F})$.

P2: For every $m \geq n$ one has $\sum_{j=n}^m a_{m,j}^2 \leq \sigma_n^2$.

Each matrix with **P1** and **P2** is an output of a greedy algorithm on some $\mathcal{F}$ e.g. $\{(f_n)_n\}$.

Matrix representation in $\mathcal{H}$

Algorithm produces $f_1, f_2, \dots$. Let f_n^*, $n = 1, 2, \dots,$ be its Gram-Schmidt orthogonalization, so $P_n f = \sum_{i=1}^n \langle f, f_i^* \rangle f_i^*$, and in particular

$$f_i = P_i f_i = \sum_{j=1}^{i} a_{i,j} f_j^*, \quad a_{i,j} = \langle f_i, f_j^* \rangle,\ j \leq i. \tag{3}$$

The lower triangular matrix $A := A(\mathcal{F}) := (a_{i,j})_{i,j=1}^{\infty}, \quad a_{i,j} = 0,\ j > i.$ contains all the information about the output of algorithm on $\mathcal{F}$. It satisfies
P1: The diagonal elements of A satisfy $|a_{n,n}| = \sigma_n := \sigma_n(\mathcal{F})$.

P2: For every $m \geq n$ one has $\sum_{j=n}^m a_{m,j}^2 \leq \sigma_n^2$.
Each matrix with **P1** and **P2** is an output of a greedy algorithm on some $\mathcal{F}$ e.g. $\{(f_n)_n\}$.

Let $d = d_m(\mathcal{F})$ and let $N > m$. Let $A_N = [a_{i,j}]_{i,j=1}^N$. We have $\det A_N = \prod_{j=1}^N \sigma_j$. Assume that there exists m dimensional space X such that each row of the matrix A_N is closer then $d = d_m$ to X in ℓ_2 norm.

$$\prod_{i=1}^N \sigma_i^2 \leq \left\{\frac{N}{m}\right\}^m \left\{\frac{N}{N-m} d^2\right\}^{N-m}. \tag{4}$$

Proof of (4): Fix an orthonormal basis for $\mathbb{R}^N$ which first spans X. Let C denotes the matrix G written in this basis. We denote by $\mathbf{c}_j$, the j-th column of C. We have

$$\begin{aligned}\prod_{i=1}^N \sigma_i^2 &= (\det A_N)^2 = (\det C)^2 \leq \prod_{j=1}^m \|\mathbf{c}_j\|_{\ell_2}^2 \cdot \prod_{j=m+1}^N \|\mathbf{c}_j\|_{\ell_2}^2 \\ &\leq \left(\frac{1}{m}\sum_{j=1}^m \|\mathbf{c}_j\|_{\ell_2}^2\right)^m \cdot \left(\frac{1}{N-m}\sum_{j=m+1}^N \|\mathbf{c}_j\|_{\ell_2}^2\right)^{N-m}\end{aligned}$$

General estimate

Theorem (DPW)

For the greedy algorithm in a Hilbert space $\mathcal{H}$ and for any compact set $\mathcal{F}$, we have the following inequalities between $\sigma_n := \sigma_n(\mathcal{F})_{\mathcal{H}}$ and $d_n := d_n(\mathcal{F})_{\mathcal{H}}$, for any $N \geq 0$, $K \geq 1$, and $1 \leq m < K$,

$$\prod_{i=1}^{K} \sigma_{N+i}^2 \leq \left\{ \frac{K}{m} \right\}^m \left\{ \frac{K}{K-m} \right\}^{K-m} \sigma_{N+1}^{2m} d_m^{2K-2m}. \tag{5}$$

Rates of decay I

(DPW)

For the greedy algorithm in a Hilbert space $\mathcal{H}$ and for any compact set $\mathcal{F}$ and $n \geq 1$, we have

$$\sigma_n(\mathcal{F}) \leq \sqrt{2} \min_{1 \leq m < n} d_m^{\frac{n-m}{n}}(\mathcal{F}). \tag{6}$$

In particular $\sigma_{2n}(\mathcal{F}) \leq \sqrt{2}\sqrt{d_n(\mathcal{F})}$, $n = 1, 2 \ldots$.

This is a delayed comparison, very general but sometimes not very efficient.

Rates of decay I

(DPW)

For the greedy algorithm in a Hilbert space $\mathcal{H}$ and for any compact set $\mathcal{F}$ and $n \geq 1$, we have

$$\sigma_n(\mathcal{F}) \leq \sqrt{2} \min_{1 \leq m < n} d_m^{\frac{n-m}{n}}(\mathcal{F}). \quad (6)$$

In particular $\sigma_{2n}(\mathcal{F}) \leq \sqrt{2}\sqrt{d_n(\mathcal{F})}$, $n = 1, 2 \ldots$.

This is a delayed comparison, very general but sometimes not very efficient.

Rates of decay II

(BCDDPW, DPW)

If $d_n(\mathcal{F}) \leq C_0 n^{-\alpha}$, $n = 1, 2, \dots,$ *then* $\sigma_n(\mathcal{F}) \leq C_1 n^{-\alpha}$, $n = 1, 2 \dots,$ *with* $C_1 := 2^{5\alpha+1} C_0$.

(DPW)

If $d_n(\mathcal{F}) \leq C_0 e^{-c_0 n^{\alpha}}$, $n = 1, 2, \dots,$ *then* $\sigma_n(\mathcal{F}) \leq \sqrt{2C_0} e^{-c_1 n^{\alpha}}$, $n = 1, 2 \dots,$ *where* $c_1 = 2^{-1-2\alpha} c_0$,

Note that for $\alpha > 1$ the estimate (2) i.e. $\sigma_n(\mathcal{K}) \leq \frac{2^{n+1}}{\sqrt{3}} d_n(\mathcal{K})$ is better–asymptotically it gives $c_1 = c_0$. For $\alpha < 1$ this is better (it preserves α). For $\alpha = 1$ the choice depends on the constants.

Rates of decay II

(BCDDPW, DPW)

If $d_n(\mathcal{F}) \leq C_0 n^{-\alpha}$, $n = 1, 2, \ldots$, *then* $\sigma_n(\mathcal{F}) \leq C_1 n^{-\alpha}$, $n = 1, 2 \ldots$, *with* $C_1 := 2^{5\alpha+1} C_0$.

(DPW)

If $d_n(\mathcal{F}) \leq C_0 e^{-c_0 n^\alpha}$, $n = 1, 2, \ldots$, *then* $\sigma_n(\mathcal{F}) \leq \sqrt{2C_0} e^{-c_1 n^\alpha}$, $n = 1, 2 \ldots$, *where* $c_1 = 2^{-1-2\alpha} c_0$,

Note that for $\alpha > 1$ the estimate (2) i.e. $\sigma_n(\mathcal{K}) \leq \frac{2^{n+1}}{\sqrt{3}} d_n(\mathcal{K})$ is better–asymptotically it gives $c_1 = c_0$. For $\alpha < 1$ this is better (it preserves α). For $\alpha = 1$ the choice depends on the constants.

Rates of decay II

(BCDDPW, DPW)

If $d_n(\mathcal{F}) \leq C_0 n^{-\alpha}$, $n = 1, 2, \ldots,$ *then* $\sigma_n(\mathcal{F}) \leq C_1 n^{-\alpha}$, $n = 1, 2 \ldots,$ *with* $C_1 := 2^{5\alpha+1} C_0$.

(DPW)

If $d_n(\mathcal{F}) \leq C_0 e^{-c_0 n^{\alpha}}$, $n = 1, 2, \ldots,$ *then* $\sigma_n(\mathcal{F}) \leq \sqrt{2C_0} e^{-c_1 n^{\alpha}}$, $n = 1, 2 \ldots,$ *where* $c_1 = 2^{-1-2\alpha} c_0$,

Note that for $\alpha > 1$ the estimate (2) i.e. $\sigma_n(\mathcal{K}) \leq \frac{2^{n+1}}{\sqrt{3}} d_n(\mathcal{K})$ is better–asymptotically it gives $c_1 = c_0$. For $\alpha < 1$ this is better (it preserves α). For $\alpha = 1$ the choice depends on the constants.

Banach spaces–Technical setup

In $\mathcal{X}$ we get vectors $f_j \in \mathcal{X}$.For each $j = 0, 1, \ldots,$ we let $\lambda_j \in \mathcal{X}^*$ be the linear functional of norm one that satisfies

$$\lambda_j(E_j) = 0, \quad \lambda_j(f_j) = \mathrm{dist}(f_j, E_j)_X = \sigma_{j+1}.$$

We associate with the greedy procedure a lower triangular matrix $A = (a_{i,j})_{i,j=0}^{\infty}$ with

$$a_{i,j} = \lambda_j(f_i).$$

So diagonal elements $a_{j,j} = \sigma_j = \mathrm{dist}(f_j, E_j)_X$.
Each entry $a_{i,j}$ satisfies

$$|a_{i,j}| = |\lambda_j(f_i)| = |\lambda_j(f_i - g)| \leq \|\lambda_j\|_{X^*}\|f_i - g\| = \|f_i - g\|, \quad j < i,$$

for every $g \in E_j$, since $\lambda_j(E_j) = 0$. Therefore we have

$$|a_{i,j}| \leq \mathrm{dist}(f_i, E_j) \leq \sigma_j, \quad j < i.$$

Banach space costs: we have an estimate for individual $A_{i,j}$'s not for their sums.

Banach spaces–Technical setup

In $\mathcal{X}$ we get vectors $f_j \in \mathcal{X}$.For each $j = 0, 1, \dots,$ we let $\lambda_j \in \mathcal{X}^*$ be the linear functional of norm one that satisfies

$$\lambda_j(E_j) = 0, \quad \lambda_j(f_j) = \mathrm{dist}(f_j, E_j)_X = \sigma_{j+1}.$$

We associate with the greedy procedure a lower triangular matrix $A = (a_{i,j})_{i,j=0}^{\infty}$ with

$$a_{i,j} = \lambda_j(f_i).$$

So diagonal elements $a_{j,j} = \sigma_j = \mathrm{dist}(f_j, E_j)_X$.
Each entry $a_{i,j}$ satisfies

$$|a_{i,j}| = |\lambda_j(f_i)| = |\lambda_j(f_i - g)| \leq \|\lambda_j\|_{X^*}\|f_i - g\| = \|f_i - g\|, \quad j < i,$$

for every $g \in E_j$, since $\lambda_j(E_j) = 0$. Therefore we have

$$|a_{i,j}| \leq \mathrm{dist}(f_i, E_j) \leq \sigma_j, \quad j < i.$$

Banach space costs: we have an estimate for individual $A_{i,j}$'s not for their sums.

Banach spaces–Technical setup

In $\mathcal{X}$ we get vectors $f_j \in \mathcal{X}$.For each $j = 0, 1, \ldots,$ we let $\lambda_j \in \mathcal{X}^*$ be the linear functional of norm one that satisfies

$$\lambda_j(E_j) = 0, \quad \lambda_j(f_j) = \mathrm{dist}(f_j, E_j)_X = \sigma_{j+1}.$$

We associate with the greedy procedure a lower triangular matrix $A = (a_{i,j})_{i,j=0}^{\infty}$ with

$$a_{i,j} = \lambda_j(f_i).$$

So diagonal elements $a_{j,j} = \sigma_j = \mathrm{dist}(f_j, E_j)_X$.
Each entry $a_{i,j}$ satisfies

$$|a_{i,j}| = |\lambda_j(f_i)| = |\lambda_j(f_i - g)| \leq \|\lambda_j\|_{X^*}\|f_i - g\| = \|f_i - g\|, \quad j < i,$$

for every $g \in E_j$, since $\lambda_j(E_j) = 0$. Therefore we have

$$|a_{i,j}| \leq \mathrm{dist}(f_i, E_j) \leq \sigma_j, \quad j < i.$$

Banach space costs: we have an estimate for individual $A_{i,j}$'s not for their sums.

Suppose that X is a Banach space. For greedy algorithm applied to a compact set $\mathcal{F}$ contained in the unit ball of $\mathcal{X}$, the following holds for $\sigma_n := \sigma_n(\mathcal{F})_X$ and $d_n := d_n(\mathcal{F})_X$, $n = 1, 2, \dots,$

(DPW)

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

$$\sigma_n \leq \sqrt{2} \min_{1 \leq m < n} n^{\frac{n-m}{2n}} \left\{ \sum_{i=1}^{n} \sigma_i^2 \right\}^{\frac{m}{2n}} d_m^{\frac{n-m}{n}}.$$

In particular $\sigma_{2\ell} \leq 2\sqrt{\ell d_\ell}$, $\ell = 1, 2 \dots$.

Estimates for Banach space II

(DPW)

If for $\alpha > 0$*, we have* $d_n \le C_0 n^{-\alpha}$, $n = 1, 2, \ldots,$ *then for any* $0 < \beta < \min\{\alpha, 1/2\}$*, we have* $\sigma_n \le C_1 n^{-\alpha+1/2+\beta}$, $n = 1, 2 \ldots,$ *with* $C_1 = C_1(\alpha, \beta, C_0)$

(DPW)

(iii) *If for* $\alpha > 0$*, we have* $d_n \le C_0 e^{-c_0 n^\alpha}$, $n = 1, 2, \ldots,$ *then* $\sigma_n < \sqrt{2C_0}\sqrt{n}e^{-c_1 n^\alpha}$, $n = 1, 2 \ldots,$ *where* $c_1 = 2^{-1-2\alpha} c_0$*. The factor* $\sqrt{n}$ *can be deleted by reducing the constant* c_1*.*

So the Banach space costs us a factor $\sqrt{n}$. This is not entirely our stupidity.

Estimates for Banach space II

(DPW)

If for $\alpha > 0$, we have $d_n \leq C_0 n^{-\alpha}$, $n = 1, 2, \ldots$, then for any $0 < \beta < \min\{\alpha, 1/2\}$, we have $\sigma_n \leq C_1 n^{-\alpha+1/2+\beta}$, $n = 1, 2 \ldots$, with $C_1 = C_1(\alpha, \beta, C_0)$

(DPW)

(iii) *If for $\alpha > 0$, we have $d_n \leq C_0 e^{-c_0 n^\alpha}$, $n = 1, 2, \ldots$, then $\sigma_n < \sqrt{2C_0}\sqrt{n} e^{-c_1 n^\alpha}$, $n = 1, 2 \ldots$, where $c_1 = 2^{-1-2\alpha} c_0$. The factor $\sqrt{n}$ can be deleted by reducing the constant c_1.*

So the Banach space costs us a factor $\sqrt{n}$. This is not entirely our stupidity.

Example

It is classical (Kashin-Gluskin) that for $\mathcal{F} = \{\pm e_j\}_{j=1}^n$ we have $d_s(\mathcal{F})_{\ell_n^\infty} \leq C\sqrt{\ln \frac{n}{s}}s^{-1/2}$ for $1 \leq s \leq n/2$. Obviously $\sigma_s(\mathcal{F}) = 1$ for $1 \leq s < n$. We use this to show

(DPW)

For any $\alpha > \frac{1}{2}$ there exists a compact set $\mathcal{F}_\alpha \subset c_0$ such that $\sigma_n(\mathcal{F}) \sim n^{-\alpha}$ and $d_n(\mathcal{F}_\alpha) \leq Cn^{-\alpha-1/2}$.

Why greedy selection is unrealistic?

Let $\mathcal{K} =: \{f_\mu \;:\; \mu \in E\}$ comes from our PDE.

- We cannot compute $\max\{\|f\| \;:\; f \in \mathcal{K}\}$ because
  1. It requires solving $D_\mu = f$ for all μ (or at least infinitely many).
  2. The solution is a function in infinite dimensional space so we cannot compute its norm.
- Next steps even more problems.
- We cannot compute f_μ exactly.

Weak greedy algorithm

For fixed $\gamma \in (0, 1]$ we define weak greedy algorithm as

1. We choose f_1 so that $\|f_1\| \geq \gamma \max\{\|f\| \ : \ f \in \mathcal{F}\}$
2. Given $f_1, \ldots, f_n$ we define $E_n = \text{span}\{f_1, \ldots, f_n\}$ and put f_{n+1} so that $\|f_{n+1}\| \geq \gamma \max\{\text{dist}(f, E_n) \ : \ f \in \mathcal{F}\}$

In the context of (1) this can be achieved cheaply.

All above results hold for weak greedy algorithm but γ must be incorporated into the constant

Weak greedy algorithm

For fixed $\gamma \in (0,1]$ we define weak greedy algorithm as

1. We choose f_1 so that $\|f_1\| \geq \gamma \max\{\|f\| \ : \ f \in \mathcal{F}\}$
2. Given $f_1, \ldots, f_n$ we define $E_n = \mathrm{span}\,\{f_1, \ldots, f_n\}$ and put f_{n+1} so that $\|f_{n+1}\| \geq \gamma \max\{\mathrm{dist}\,(f, E_n) \ : \ f \in \mathcal{F}\}$

In the context of (1) this can be achieved cheaply.

All above results hold for weak greedy algorithm but γ must be incorporated into the constant

Realistic scheme

We have one more problem: when we compute f_j we obtain some errors. We have $\gamma \in (0,1]$ and accuracy $\epsilon > 0$

1. At the first step we determine $f_1 \in \mathcal{F}$ such that $\|f_1\| \geq \gamma \sup_{f\in\mathcal{F}} \|f\|$. However, from computation we receive the noisy version $\hat{f}_1$ which is not even necessarily in $\mathcal{F}$. All we know about $\hat{f}_1$ is that

$$\|f_1 - \hat{f}_1\| \leq \varepsilon.$$

2. Given $\hat{f}_1, \dots, \hat{f}_n$ we define $\hat{E}_n = \mathrm{span}\{\hat{f}_1, \dots, \hat{f}_n\}$ and determine $f_{n+1} \in \mathcal{F}$ such that $\mathrm{dist}(f_{n+1}, \hat{E}_n) \geq \gamma \sup_{f\in\mathcal{F}} \mathrm{dist}(f, \hat{E}_n)$. Rather than f_{n+1} we receive the noisy version $\hat{f}_{n+1}$ for which we only know $\|f_{n+1} - \hat{f}_{n+1}\| \leq \varepsilon$.

Let us define

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

which is the performance of this realistic scheme on $\mathcal{F}$.

Performance of the realistic scheme

Theorem (BCDDPW)

Suppose we are given $\gamma \in (0,1]$ *and* $\epsilon > 0$. *Suppose that*

$$d_n(\mathcal{F}) \leq M n^{-\alpha}, \quad n \geq 0, \tag{7}$$

for some M *and* $\alpha > 0$. *Then,*

$$\hat{\sigma}_n(\mathcal{F}) \leq C \max\{M n^{-\alpha}, \varepsilon\}, \qquad n \geq 0, \tag{8}$$

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

Happy Birthday Freddy