Lecture 1: Minimax Polynomial Approximation

Pr. Ismail Merabet

Univ. of K-M-Ouargla

October 12, 2024

Contents

Introduction

We are interested in the approximaton of a given function f by a function belonging to some particular subspace S.In particular, we will be interested in the case $S = \mathcal{P}_n(I)$.

Introduction

We are interested in the approximaton of a given function f by a function belonging to some particular subspace S.In particular, we will be interested in the case $S = \mathcal{P}_n(I)$.
We shall be interested in two sorts of questions:

Introduction

We are interested in the approximaton of a given function f by a function belonging to some particular subspace S.In particular, we will be interested in the case $S = \mathcal{P}_n(I)$.
We shall be interested in two sorts of questions:

- How good is the best approximation?

Introduction

We are interested in the approximaton of a given function f by a function belonging to some particular subspace S.In particular, we will be interested in the case $S = \mathcal{P}_n(I)$.
We shall be interested in two sorts of questions:

- How good is the best approximation?
- How good are various computable approximation procedures?

Introduction

We are interested in the approximaton of a given function f by a function belonging to some particular subspace S.In particular, we will be interested in the case $S = \mathcal{P}_n(I)$.

We shall be interested in two sorts of questions:

- How good is the best approximation?
- How good are various computable approximation procedures?

We shall always use a norm to specify the goodness of the approximation.

Introduction

We are interested in the approximaton of a given function f by a function belonging to some particular subspace S.In particular, we will be interested in the case $S = \mathcal{P}_n(I)$.

We shall be interested in two sorts of questions:

- How good is the best approximation?
- How good are various computable approximation procedures?

We shall always use a norm to specify the goodness of the approximation.

We shall take up the first question, the theory of best approximation, first.

Thus we want to know about :

Introduction

We are interested in the approximaton of a given function f by a function belonging to some particular subspace S.In particular, we will be interested in the case $S = \mathcal{P}_n(I)$.

We shall be interested in two sorts of questions:

- How good is the best approximation?
- How good are various computable approximation procedures?

We shall always use a norm to specify the goodness of the approximation.

We shall take up the first question, the theory of best approximation, first.

Thus we want to know about :

$$\inf_{p \in P} \|f - p\|$$

for some specified norm.

Various questions come immediately to mind:

Various questions come immediately to mind:

- Does there exist $p \in P$ minimizing $\|f - p\|$?

Various questions come immediately to mind:

- Does there exist $p \in P$ minimizing $\|f - p\|$?
- Could there exist more than one minimizer?

Various questions come immediately to mind:

- Does there exist $p \in P$ minimizing $\|f - p\|$?
- Could there exist more than one minimizer?
- Can the (or a) minimizer be computed?

Various questions come immediately to mind:

- Does there exist $p \in P$ minimizing $\|f - p\|$?
- Could there exist more than one minimizer?
- Can the (or a) minimizer be computed?
- What can we say about the error?

Various questions come immediately to mind:

- Does there exist $p \in P$ minimizing $\|f - p\|$?
- Could there exist more than one minimizer?
- Can the (or a) minimizer be computed?
- What can we say about the error?

The answer to the first question is affirmative under quite weak hypotheses. To see this, we first prove a simple lemma.

Various questions come immediately to mind:

- Does there exist $p \in P$ minimizing $\|f - p\|$?
- Could there exist more than one minimizer?
- Can the (or a) minimizer be computed?
- What can we say about the error?

The answer to the first question is affirmative under quite weak hypotheses. To see this, we first prove a simple lemma.

Lemma

Let there be given a normed vector space X and $n+1$ elements $f_0, \ldots, f_n$ of X. Then the function $\phi : \mathbb{R}^n \to \mathbb{R}$ given by :

$$\phi(\boldsymbol{a}) = \left\| f_0 - \sum_{i=1}^{n} a_i f_i \right\|$$

is continuous.

Theorem

Let there be given a normed vector space X and a finite dimensional vector subspace P.

Theorem

Let there be given a normed vector space X and a finite dimensional vector subspace P. Then for any $f \in X$ there exists $p \in P$ minimizing $\|f - p\|$.

Theorem

Let there be given a normed vector space X and a finite dimensional vector subspace P. Then for any $f \in X$ there exists $p \in P$ minimizing $\|f - p\|$.

Proof.

Let $f_1, \ldots, f_n$ be a basis for P. The map $\boldsymbol{a} \mapsto \|\sum_{i=1}^{n} a_i f_i\|$ is then a norm on $\mathbb{R}^n$. Hence it is equivalent to any other norm, and so the set

$$S = \left\{ \boldsymbol{a} \in \mathbb{R}^n \mid \left\| \sum a_i f^i \right\| \leq 2\|f\| \right\},$$

is closed and bounded. We wish to show that the function $\phi : \mathbb{R}^n \to \mathbb{R}, \phi(\boldsymbol{a}) = \|f - \sum a_i f_i\|$ attains its minimum on $\mathbb{R}^n$. By the lemma this is a continuous function, so it certainly attains a minimum on S, say at $\boldsymbol{a}_0$. But if $\boldsymbol{a} \in \mathbb{R}^n \setminus S$, then

$$\phi(\boldsymbol{a}) \geq \left\| \sum a_i f_i \right\| - \|f\| > \|f\| = \phi(\mathbf{0}) \geq \phi(\boldsymbol{a}_0)$$

This shows that $\boldsymbol{a}_0$ is a global minimizer.

Defintion

A norm is called strictly convex if its unit ball is strictly convex. That is, if $\|f\| = \|g\| = 1$, $f \neq g$, and $0 < \theta < 1$ implies that $\|\theta f + (1-\theta)g\| < 1$.

Example

The L^p norm is strictly convex for $1 < p < \infty$, but not for $p = 1$ or ∞.

Theorem

Let X be a strictly convex normed vector space, P a subspace, $f \in X$, and suppose that p and q are both best approximations of f in P. Then $p = q$.

Proof.

By hypothesis $\|f - p\| = \|f - q\| = \inf_{r \in P} \|f - r\|$. By strict convexity, if $p \neq q$, which is impossible.

$$\|f - (p+q)/2\| = \|(f-p)/2 + (f-q)/2\| < \inf_{r \in P} \|f - r\|$$

We shall now focus on the case of best approximation by polynomials of degree at most n measured in the L^∞ norm (minimax approximation).

We shall now focus on the case of best approximation by polynomials of degree at most n measured in the L^∞ norm (minimax approximation). We first show that arbitrarily good approximation is possible if the degree is high enough.

We shall now focus on the case of best approximation by polynomials of degree at most n measured in the L^∞ norm (minimax approximation). We first show that arbitrarily good approximation is possible if the degree is high enough.

(Weierstrass Approximation Theorem)

Let $f \in C(I)$ and $\epsilon > 0$. Then there exists a polynomial p such that $\|f - p\|_\infty \leq \epsilon$

We shall give a constructive proof due to S. Bernstein. But let us first recall the definition of Bernstein's polynomial and some of its properties.

We shall now focus on the case of best approximation by polynomials of degree at most n measured in the L^∞ norm (minimax approximation). We first show that arbitrarily good approximation is possible if the degree is high enough.

(Weierstrass Approximation Theorem)

Let $f \in C(I)$ and $\epsilon > 0$. Then there exists a polynomial p such that $\|f - p\|_\infty \leq \epsilon$

We shall give a constructive proof due to S. Bernstein. But let us first recall the definition of Bernstein's polynomial and some of its properties.

Defnition

For $f \in C(I), n = 1, 2, \ldots,$ define $B_n f \in \mathcal{P}_n(I)$ by

$$B_n f(x) = \sum_{k=0}^{n} f\left(\frac{k}{n}\right) C_k^n x^k (1-x)^{n-k}$$

Properties of Bernstein's poly

Lemma

1. $B_n(1) = 1, \quad$ (i.e, for $f(x) = 1$).

Properties of Bernstein's poly

Lemma

1. $B_n(1) = 1, \quad$ (i.e, for $f(x) = 1$).
2. $B_n(x) = x, \quad$ (i.e, for $f(x) = x$.

Properties of Bernstein's poly

Lemma

1. $B_n(1) = 1, \quad$ (i.e, for $f(x) = 1$).
2. $B_n(x) = x, \quad$ (i.e, for $f(x) = x$.
3. $B_N(x^2) = \frac{n-1}{n}x^2 + \frac{x}{n}, \quad$ (i.e, for $f(x) = x^2$.

Properties of Bernstein's poly

Lemma

1. $B_n(1) = 1, \quad$ (i.e, for $f(x) = 1$).
2. $B_n(x) = x, \quad$ (i.e, for $f(x) = x$.
3. $B_N(x^2) = \frac{n-1}{n}x^2 + \frac{x}{n}, \quad$ (i.e, for $f(x) = x^2$.

Proof

To prove these identities, first, from the binomial theorem,

$$B_n(1) = \sum_{k=0}^{n} C_k^n x^k (1-x)^{n-k} = [x + (1-x)]^n = 1.$$

Properties of Bernstein's poly

Lemma

1. $B_n(1) = 1, \quad$ (i.e, for $f(x) = 1$).
2. $B_n(x) = x, \quad$ (i.e, for $f(x) = x$.
3. $B_N(x^2) = \frac{n-1}{n}x^2 + \frac{x}{n}, \quad$ (i.e, for $f(x) = x^2$.

Proof

To prove these identities, first, from the binomial theorem,

$$B_n(1) = \sum_{k=0}^{n} C_k^n x^k (1-x)^{n-k} = [x + (1-x)]^n = 1.$$

Note that

$$\frac{d}{dp}\left(\sum_{k=0}^{n} C_k^n p^k q^{n-k}\right) = \frac{d}{dp}\left((p+q)^n\right) = n(p+q)^{n-1}.$$

Thus

$$\sum_{k=0}^{n} C_k^n \frac{k}{n} p^k q^{n-k} = (p+q)^{n-1} p.$$

Replacing p by x and q by $1-x$ in the above expression, we have identity (2).

Thus

$$\sum_{k=0}^{n} C_k^n \frac{k}{n} p^k q^{n-k} = (p+q)^{n-1} p.$$

Replacing p by x and q by $1-x$ in the above expression, we have identity (2).Now, differentiating this expression with respect to p three more times and each time multiplying both sides of the result by $\frac{p}{n}$, we have the following

Thus

$$\sum_{k=0}^{n} C_k^n \frac{k}{n} p^k q^{n-k} = (p+q)^{n-1} p.$$

Replacing p by x and q by $1-x$ in the above expression, we have identity (2).Now, differentiating this expression with respect to p three more times and each time multiplying both sides of the result by $\frac{p}{n}$, we have the following

$$\sum_{k=0}^{n} \left(\begin{array}{c} n \\ k \end{array} \right) \frac{k^2}{n^2} p^k q^{n-k} = \frac{(n-1)(p+q)^{n-2}}{n} p^2 + \frac{(p+q)^{n-1}}{n} p$$

Proof of Weirstrass Theorem

By Hein's theorem :

$$\forall \varepsilon > 0, \exists \delta > 0, \quad \text{s.t} \quad \forall x, x'; |x - x'| < \delta \implies |f(x) - f(x')| < \frac{\varepsilon}{2}$$

Proof of Weirstrass Theorem

By Hein's theorem :

$$\forall \varepsilon > 0, \exists \delta > 0, \quad \text{s.t} \quad \forall x, x'; |x - x'| < \delta \implies |f(x) - f(x')| < \frac{\varepsilon}{2}$$

From property 1 we have:

$$|f(x) - B_n(f)(x)| = \left| \sum_{k=0}^{n} \left(f(x) - f\left(\frac{k}{n}\right) \right) C_k^n x^k (1-x)^{n-k} \right|$$

Proof of Weirstrass Theorem

By Hein's theorem :

$$\forall \varepsilon > 0, \exists \delta > 0, \quad \text{s.t} \quad \forall x, x'; |x - x'| < \delta \implies |f(x) - f(x')| < \frac{\varepsilon}{2}$$

From property 1 we have:

$$|f(x) - B_n(f)(x)| \;=\; \left| \sum_{k=0}^{n} \left(f(x) - f\left(\frac{k}{n}\right) \right) C_k^n x^k (1-x)^{n-k} \right|$$

To estimate this last sum, we separate the terms into two sums $\sum^{(1)}$ and $\sum^{(2)}$, those where $\left|\frac{k}{n} - x\right|$ is less than a given positive δ and the remaining terms, those for which $\delta \leq \left|\frac{k}{n} - x\right|$.

Suppose that x is a point of continuity of f.

Suppose that x is a point of continuity of f.Then for any $\varepsilon > 0$, there is a positive δ such that

$$\left| f\left(\frac{k}{n}\right) - f(x) \right| < \frac{\varepsilon}{2} \text{ when } \left| \frac{k}{n} - x \right| < \delta.$$

For the first sum,

$$\begin{aligned} \sum^{(1)} C_k^n x^k (1-x)^{n-k} \left| f\left(\frac{k}{n}\right) - f(x) \right| &< \sum^{(1)} C_k^n x^k (1-x)^{n-k} \frac{\varepsilon}{2} \\ &\le \frac{\varepsilon}{2} \sum_{k=0}^{n} C_k^n x^k (1-x)^{n-k} = \frac{\varepsilon}{2} \end{aligned}$$

For the remaining terms,

For the remaining terms,we have $\delta^2 \le \left|\frac{k}{n} - x\right|^2$,

$$\begin{aligned}
\delta^2 \sum^{(2)} C_k^n x^k (1-x)^{n-k} \left| f\left(\frac{k}{n}\right) - f(x) \right| \le & \sum^{(2)} C_k^n \left(\frac{k}{n} - x\right)^2 x^k (1-x)^{n-k} \left| f\left(\frac{k}{n}\right) - \right. \\
\le & \sum^{(2)} C_k^n \left(\frac{k}{n} - x\right)^2 x^k (1-x)^{n-k} 2M \\
\le & 2M \sum_{k=0}^{n} C_k^n \left(\frac{k}{n} - x\right)^2 x^k (1-x)^{n-k} \\
= & 2M \frac{x(1-x)}{n} \\
\le & \frac{2M}{n}.
\end{aligned}$$

For the remaining terms,we have $\delta^2 \leq \left|\frac{k}{n} - x\right|^2$,

$$\begin{aligned}
\delta^2 \sum^{(2)} C_k^n x^k (1-x)^{n-k} \left|f\left(\frac{k}{n}\right) - f(x)\right| \leq & \sum^{(2)} C_k^n \left(\frac{k}{n} - x\right)^2 x^k (1-x)^{n-k} \left|f\left(\frac{k}{n}\right) - \right. \\
\leq & \sum^{(2)} C_k^n \left(\frac{k}{n} - x\right)^2 x^k (1-x)^{n-k} 2M \\
\leq & 2M \sum_{k=0}^{n} C_k^n \left(\frac{k}{n} - x\right)^2 x^k (1-x)^{n-k} \\
= & 2M \frac{x(1-x)}{n} \\
\leq & \frac{2M}{n}.
\end{aligned}$$

Thus

$$\sum^{(2)} C_k^n x^k (1-x)^{n-k} \left|f\left(\frac{k}{n}\right) - f(x)\right| \leq \frac{2M}{\delta^2 n}.$$

For this δ, we can choose n_0 large enough so that, when $n \geq n_0$, $\frac{2M}{\delta^2 n} < \frac{\varepsilon}{2}$.

For this δ, we can choose n_0 large enough so that, when $n \geq n_0$, $\frac{2M}{\delta^2 n} < \frac{\varepsilon}{2}$.
For such an n and the given x

For this δ, we can choose n_0 large enough so that, when $n \geq n_0$, $\frac{2M}{\delta^2 n} < \frac{\varepsilon}{2}$. For such an n and the given x

$$|B_n(f)(x) - f(x)| \leq \overset{(1)}{\sum} + \overset{(2)}{\sum} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon$$

Hence $B_n(f)(x) \to f(x)$ as $n \to \infty$ for each point x of continuity of the function f.

Remarks

From a practical point of view the Bernstein polynomials yield :

Remarks

From a practical point of view the Bernstein polynomials yield :

- an approximation procedure which is very robust but very slow. By robust we refer to the fact that the procedure convergence for any continuous function.

Remarks

From a practical point of view the Bernstein polynomials yield :

- an approximation procedure which is very robust but very slow. By robust we refer to the fact that the procedure convergence for any continuous function.
- Moreover if the function is C^1 then not only does $B_n f$ converge uniformly to f, but $(B_n f)'$ converges uniformly to f' (i.e., we have convergence of $B_n f$ in $C^1(I)$, and similarly if f admits more continuous derivatives.

Remarks

From a practical point of view the Bernstein polynomials yield :

- an approximation procedure which is very robust but very slow. By robust we refer to the fact that the procedure convergence for any continuous function.
- Moreover if the function is C^1 then not only does $B_n f$ converge uniformly to f, but $(B_n f)'$ converges uniformly to f' (i.e., we have convergence of $B_n f$ in $C^1(I)$, and similarly if f admits more continuous derivatives.
- However, even for very nice functions the convergence is rather slow.

Example

Take as simple a function as $f(x) = x^2$, we saw that

$$\|f - B_n f\| = O(1/n).$$

In fact, refining the argument of the proof, one can show that this same linear rate of convergence holds for all C^2 functions f :

$$\|f - B_n f\| \leq \frac{1}{8n} \|f''\|, \quad f \in C^2(I).$$

this bound holds with equality for $f(x) = x^2$, and so cannot be improved. This slow rate of convergence makes the Bernstein polynomials impractical for most applications.