GREEDY APPROXIMATION AND APPLICATIONS OF GREEDY APPROXIMATION IN DISCREPANCY

VLADIMIR TEMLYAKOV

Abstract. The following notes were taken by Miss J.L. Nelson during Prof. Temlyakov's lecture for the Spring 2008 High Dimensional Approximation Seminar. Please note that the fragments of sentences are the result of these notes being not yet correlated with the transcript of the talk. A final version should be available by the end of June 2008.

Greedy approximation arises in a number of settings including both finite and infinite-dimensional spaces such as Hilbert spaces and Banach spaces.

Since this seminar series is focussed on high dimensional concepts, I have chosen to address the idea of discrepancy to illustrate how greedy approximation techniques, which can be developed in very general settings, can be applied to high-dimensional problems.

We will begin with the concept of discrepancy, which originated in number theory as the problem of finding a set of well-distributed points in a particular domain.

Let's consider, then, the following simple domain, the unit cube:

$$\Omega_d := [0, 1]^d$$

Over this domain we will consider a set of points, each with d coordinates

In two dimension

$\xi = (\xi^1, ..\xi^m)\xi^j \in \Omega_d$

Now the question becomes, "How can we measure how well these points are distributed in the unit cube?"

There are several ways to do this, but the one which arises from discrepancy is as follows.

First we define a function

$$B(x, y) := \prod \chi_[0, x_i](y_i)$$

$x = (x_1, ...x_d) \in \Omega_d$

$y = (y_1, ..., y_d) \in \Omega_d$

$$D(\xi, m, d)_p := \| \int_\Omega B(\cdot, y)dy - \frac{1}{m} \sum_{j=1}^{m} B(\cdot, \xi^j)\|_p$$

Our goal is to study this quantity

$$D(m, d)_p := \inf_\xi D(\xi, m, d)_p$$

...It is known, like..

Date: March 19, 2008.

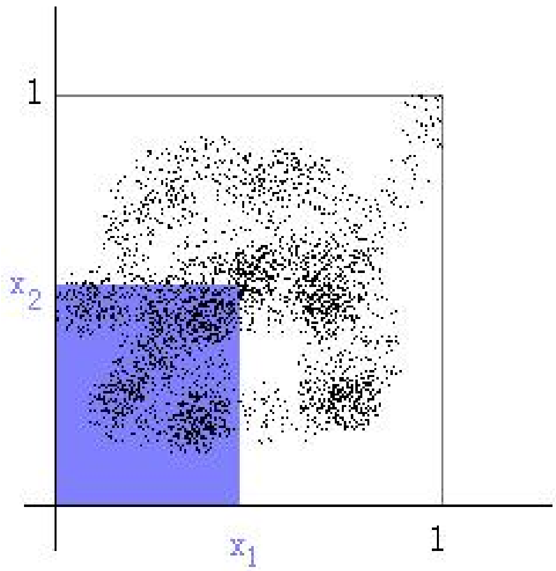

FIGURE 1. Discrepancy thingamaggly

right away before we are writing any results on this... greedy approximation .. what kind of nonlinear approximation problem do we have here?..

$$J(x) := \in_{\Omega_d} B(x,y)dy$$

we approximation this specific function ..

if you write this dictionary d

$$\mathcal{D} = \{B(\cdot,y), y \in \Omega_d\}$$

now these functions .. on x.. so it is a dictionary.. best m-term approximation of a specific function.. what is typical .. do in approximation theory..

$$\mathcal{D} = \{g\}\|g\| \le 1$$

$$\sigma_m(f,\mathcal{D}) := \inf_{\{c_j\},\{g_j\}\subset\mathcal{D}} \|\sum_{j=1}^{m} c_j g_j\|$$

$$\|B(\cdot,y)\|_p \le 1$$

$1 < p < \infty$

$$D(m,d)_p \asymp \frac{(\log m)^{\frac{d-1}{2}}}{m}$$

$\overset{d,p}{\asymp}$

(K. Roth, K .Frolov , 1954-1980)

the constants above depend on d and p

amazingly good order because we have just log .. you cannot apply this result...

(S. Heinrich, E. Novak, G. Wasilkowski, H. Wozniakowski 2001) proved the following estimate

$$D(m,d)_\infty \leq Cd^{1/2}m^{-1/2}$$

the advantage of this one is a very good dependence on d

so this is a kind of trade-off.. the proof was probabilistic.. in order to illustrate a bit.. why this $\sqrt(m)$ appears

Remark 0.1. Monte Carlo method

$$\int_\Omega f(x)dx - \frac{1}{N}\sum_{j=1}^N f(\xi^j)$$

this is a function which depends on all these points

$$\begin{aligned}\int_{\Omega^N}(\int_\Omega f(x)dx - \frac{1}{N}\sum_{j=1}^N f(\xi^j))^2 d\xi^1 ... d\xi^N &= \int_{\Omega^N}(A^2 - \frac{2A}{N}\sum_{j=1} f(\xi^j) + \frac{1}{N^2}(\sum_{j=1}^N f(\xi^j)^2) d\xi^1 \cdots d\xi^N \\ &= A^2 - \frac{2A^2}{N}N + \frac{1}{N^2}(N\int_\Omega f(x)^2 dx + (N^2 - N)A^2) \\ &= -\frac{A^2}{N} + \frac{\|f\|_2^2}{N} \leq \frac{\|f\|_2^2}{N}\end{aligned}$$

it is an expectation .. of this quantity .. if we .. with the same probability.. this is exactly expectation.. if you take one specific function.. with high probability.. square of the error.. so the

so with high prob..

error $\leq \frac{\|f\|_2}{\sqrt{N}}$;

concentration measure inequalities and so on.. I believe these guys used this .. how you

they used this probabilistic method .. prove that these points exist.. don't even know that these points exist.. in any practical problem.. constructive way to build these points..

too general and too .. optimistic. can be considered as a deterministic.. to build these points..

what should be the theorem here in .. which could be applied in our situation..

what is the procedure? I will try to give you different .. but the one with is applied here is called incremental ..

1. Incremental Algorithm with Schedule ε, $\varepsilon_k > 0$

but this one is not the simplest one.. first I will discuss .. some ... greedy approx. may help in problems like this..

Let $f \in A_1(D) := \overline{\text{conv}(\mathcal{D})}$, so set $f_0 := f$ and $G_0 := 0$

inductively, assume that we have f_{m-1}, G_{m-1}

step 1: Let

$$\phi_m \in \mathcal{D} : F_{f_{m-1}}(\phi_m - f) \geq -\varepsilon_m$$

step 2:

$$G_m := (1 - \frac{1}{m})G_{m-1} + \frac{1}{m}\phi_m$$

this is called relaxation
it is clear

$$G_m = \frac{1}{m}\sum_{j=1}^{m}\phi_j$$

step 3: $f_m := f - G_m$.

Let me explain the concepts which are involved here.
First of all the norming functional

$$h \in X, F_h \in X^*, with the follwoing prop. \|F_h\|_{X^*} = 1, F_H(h) = \|h\|$$

In general this functional is not unique.. in that case it is always unique.. we need to associate this functional for this ..
why (step 1)

Lemma 1.1. *For all bounded $F \in X^*$ we have*

$$\sup_{g\in\mathcal{D}} F(g) = \sup_{h\in A_1(\mathcal{D})} F(h).$$

If so, if we take sup of step 1 then ..

$$\sup_{g\in\mathcal{D}} F(g) = \sup_{h\in A_1(\mathcal{D})} F(h) \geq F(f)$$

..so it exists..
..with $\varepsilon_k > 0$.. in this setting .. always exists

Theorem 1.2. *Let X be a uniformly smooth Banach space with modulus of smoothness $\rho(u) \leq \gamma u^q, 1 < q \leq 2$. Define*

$$\varepsilon_n := K_1\gamma^{1/q}n^{-1/p}$$

where

$$\frac{1}{q} + \frac{1}{p} = 1 \ (i.e.\ p := \frac{q}{q-1}).$$

Then for any $f \in A_1(\mathcal{D})$,

$$\|f_m\| \leq c(K_1)\gamma^{1/q}m^{-1/p}.$$

.. we apply this procedure and we're done..

$$\rho(u) := \sup_{\|x\|=\|y\|=1}((\frac{1}{2}\|x - uy\| + \|x - uy\|) - 1)$$

Uniformly smooth Banach Space: $\lim_{n\to 0}\frac{\rho(u)}{u}$
L_p
if $1 < p < \infty$

$$\rho(u) = \begin{cases} \frac{u^p}{p}, & \text{if } 1 \leq p \leq 2 \\ \frac{p-1}{2}u^2, & \text{if } 2 \leq p < \infty \end{cases}$$

If we apply this theorem with this definition.

All coefficients are $1/n$ .. belongs to the convex hull... easy to see from here.. think of this function $(D(\xi, m, d)_p)$

It is clear it is a limit

We can write

$J \in A_1(\mathcal{B})$

Now what we get

$\|f_m\|_p \leq C$..

It is clear that we can restrict ourselves to $2 \leq p < \infty$

$\|f_m\|_p \leq Cp^{1/2}m^{-1/2}$

This constant depends on neither d nor p.

This space L_∞ is not uniformly smooth, so this technique.. compare this with $D_(m, d)_\infty (above)$

Some remarks.. ... Dahmen:

Temlyakov: ... that's it. ...

Now let me formulate one or two problems.. the idea is to build this m-term approx using.. we are looking for ϕ_m .. put even 0 here (where ever "here" is)

–

Hilbert space H

$$\mathcal{D} = \{g\} \|g\| = 1$$

Pure Greedy Algorithm otherwise you cannot approximate.. not necessary the one form the convex hull...

Initialization:: $f_0 := f, G_0 := 0$

Step 1::

$$\phi_m \in \mathcal{D} : |\langle f_{m-1}, \phi_m \rangle| = \sup_{g \in \mathcal{D}} |\langle f_{m-1}, g \rangle|$$

$$F_h(w) := \langle \frac{h}{\|h\|, w} \rangle$$

Step 2::

$$f_m := f_{m-1} - \langle f_{m-1}, \phi_m \rangle \phi_m$$

(L. Jones 1987, conj. by Huber) For all $f \in H$, $f_m \to 0$.

We are interested in the rate of approximation .. This example shows that.. natural classes to look at..

In this case

$$f \in A_1(\mathcal{D}) := \overline{\mathrm{conv}(D^{\pm})}$$

$$\|f_m\| \leq m^{-1/6}, m^{-11/62}$$

Sil'michenko

$$\|f_m\| \leq m^{-0.181..}$$

have this estimate kind of universal for any.. fix .. build this k Hilbert space..

$$\|f_m\| \geq m^{-0.27}, m^{-0.189}$$

the technique for upper estimates and lower estimates end up with ..

Dahmen: so, pure greed is not good.

Questions

Dahmen: this particular version looks more like relaxed greedy..

Temlyakov: yes, you're absolutely right.. there are different versions of this .. we build like this, all the time.. we also choose this ϕ_m .. sort of similar..

Dahmen: peak functional is just inner product..

Temlyakov: ..concept of weak greedy algorithms.. weak version of .. small, even going to zero.. co-convex..

Dahmen: so this is a strange phenomenon.. as cheap as the pure is, much cheaper than doing the ortho.. performs as well as .. very structured dictionaries..

Temlyakov: as to relaxation.. clarify.. in Banach spaces, there is a version called "free" relaxation.. instead of this fixed ..

you optimize over two parameters.. rates of approximation.. pretty good substitute in a sense.. very good!

Bennet: Anything in $p = 1$?

Temlyakov:.. my personal opinion is.. is considered the most difficult.. $p = 1$ is more difficult..

Pavel: return back to ..

Temlyakov: ..

University of South Carolina
E-mail address: temlyak@math.sc.edu