Draft copy 5/12/2008

Rearrangement of Maximal Functions Math 565 Presentation for Professor Marie Christina Pereyra

C. James Elliott[1][2]

Student, Mathematics and Statistics Dept, UNM

cje@unm.edu

1. The Flame of Decreasing Rearrangement Operators

The topic of maximal operators appeals to me because of its Algebric character and the secret mantra Decreasing Rearrangements. The maximal character of the Hardy-Littlewood functional we describe here also reminds me of information entropy and the Boltzmann's H theorem in plasma physics. Morally everything done before was done in a fog and once you look at things through the eyes of Maximal Operators and Decreasing Rearrangements, you will be totally hooked, drums will beat, whistles will blow and the devine light from above will shine upon you. You will be well pleased to be able to site Hardy and Littlewood, *A maximal theorem with function theoretical applications* Acta Math 54 (1930) pages 81-116 and download it free of charge from the campus web. We will also give a geometric proof of H-L Theorem I in terms of a barycentric decomposition well known in algebraic geometry and computer science (permutaion generation) and an argument from a weight diagram that reduces Theorem II to Theorem I. In addition we hope to clarify equivalence relationships amoung maximal operators so that utilization of them or their proxies will be natural.

2. Part I: A Retold Story: Decreasing Rearrangements are Cricket

We know that Hardy and Littlewood described maximal operators in a one sided way in one dimension, and they described their idea in terms of the game of cricket. Remarkedly, the same paper also introduces decreasing rearrangements. Theorem I states what they call a well known result: Suppose a cricket batsman in a given season has played a number of

[1]Laboratory Affiliate, C-PCS, Los Alamos National Laboratory, Los Alamos NM, 87545

[2]Non-degree Graduate Student, University of New Mexico, Albuquerque NM

games of cricket. The nefarious manager as managers do takes the average after each game much as one does with Cesaro sums and assigns a satisfaction score for each game based only on an increasing function of the average over all the preceding games. He wants a metric on which to base your pay raise he calls his total satisfaction with you. Your satisfaction at the end of the last game of the season doesn't cut it for him even though it measures your average performance. Your manager believes that if you do well at first your early scores will make a lasting impression and attract more fans to the game. His metric for your total performance at season's end is based on the sum of the satisfactions at the end of all the games. The more fans you satisfy early in the season the more they will tell their friends and the higher will be your pay raise. The satisfaction of your effort for each prior game will be remembered at the end of a susceeding game for the purposes of determining the satisfaction then, and these end of game satisfactions including a non-linear increasing bonus are totaled at the end of the season giving the manager's total satisfaction with you.

In Fig.(1) We show the relationship of the game averages on an affine (Singular) 3-Simplex correspond to barycenter images. It is an example of a four game season and it shows how to plot the relevant data on a 3-simplex. One may think of the grey barycenter of the three games as the overall average for a three-game season. The fact that the overall average wasn't good enough for the manager, led to the model used in Theorem I the Accumulated Memory Model shown in Fig.(2). Quadrant I has the game averages. Quadrant II shows the weights applied to arrive at the accumulated averages in Quadrant III. Each accumulated average can be thought of as an actual compensation for the following year with possible bonus. Without bonuses, the sum of the accumulated averages determine next's year's raise. With bonuses, it is rather the sum of the four blue ordinates that determine next year's raise. The bonus curve is strictly increasing in our case for a realistic game, but can be allowed to be but non-decreasing and non-zero in the Hardy Littlewood paper. If you add their caveats, it makes no difference.

The most basic lemma is this: Given a set of decreasing weights and game scores that can be assigned one-one in any order to the weights, the maximum average results in the game scores in a decreasing rearrangement. Every student knows: do best on tests where the weight is highest. In Fig. 2 the sum of weights is clearly decreasing and this arrangement of weights induces a corresponding arrangement of scores.

Fig. 3 Shows some basic properties of barycentric coordinates. Note that here the axes in the source space are equal length where as they are not in the image space. Fig.4 Shows the Fans' Memory and Satisfaction Model with the new weights and with choices.

Theorem II is a maximal version of Theorem I and is the essence of the paper according to Hardy and Littlewood, but it is still easy to follow the claim although more difficult to believe at first sight. Fig. 5 Shows the Fans' Model corresponding to Theorme II for having optimum pay raise reduces to the same four cases as before. The amount of the raise is different but the decreasing rearrangement holds. Your team has convinced the manager that fans can't remember all the way to the beginning of the season. They still remember average performance over games, and like to discuss best averages in the pub at the end of each game. They could remember your average performance over any of the last n games, but that's it. The one they choose to remember and give to the manager after each game is perhaps back to 5 games ago or back to 2 games ago, back to n games ago whenever your best average over the last n games occurred. What does the manager do with this maximal average? Just as before he puts each one into a pot and adds them all up at the end of the season after applying his non-linear increasing function to give spectular-performance bonuses and thereby determines your pay raise. Believe it or not if you could rearrange the games you played in any order your raise would be highest again if your performance over individual games occurred in decreasing order. That's the theorem.

To understand the proof, first note that all of the choices of the first model are present in the second. In G4 the choices correspond to face maps: first deleting the first vertex, then the second, and then the third. Now suppose vertex1 has the highest game average, vertex2 the next highest ...and the last vertex has the lowest average score. So each face map represents a bad choice. What are you left with? The choice for Gi contains vertex 1 in all cases. Lo and behold, we are back to the same choices as in the previous model. The players wished that the manager would let them rearrange their scores, but he wouldn't do it. He wanted the difference in raises to provide an incentive. "Anyway, he told the players,"Just do your best early in the season, and with decreasing averages it won't make any difference what scheme I use. Don't get the idea though that you should do less well on the second game after you have played the first. The second game score is the second most important score for you to keep high if you want a good raise."

Why did Hardy and Littlewood not report a geometric result? My guess is that they were in a Cricket Club after a big Cambridge vs Oxford game. "I got this great result in barycenter coordinates one said." But just then the Maitre de came by and said, "Gentlemen, may I remind you this is a Cricket Club. (pause) What you are discussing, is not Cricket!" So the mathematical results were translated into game averages and satisfaction and everyone was happy. In fact or fiction Hardy and Littlewood liked that version better so you have it.

It should be no surprise that with the initially astonishing results of Theorem II that decreasing rearrangements are an important way to classify functions. You construct your blocks corresponding to the optimum choice weights and remember only these maximal averages of the function. Then you raise that value to the $p'th$ power or some other increasing function, Φ and add them all up to see how well the function performed at that location. If you only need a bound for this maximal sum you call normal performance or norm for short when you take its p'th root, you can get the upper bound from the decreasing rearranged function's maximal function by computing its norm. This works for all p and also suggests Orlicz types of Banach spaces L^Φ possibly after generalizing the Luxemburg norm (Duoandikoetxea). Exercise: Determine whether Hardy and Littlewood's paper came before that of Orlicz; did Luxemberg's paper come before Orlicz's? If not so, how did Orlicz define his norm? Is there only a problem in defining norms with $f \log^+ f$? Is Orlicz's norm essentially that occuring in Hardy and Littlewood's paper? Does Orlicz generalize Luxemburg?

3. Part II: Using Hardy and Littlewood's Results

The above dialogue suggests the following continuous version of that discrete result as a conjecture:

$$||M^- f(x)||_p \leq ||M^- f^*||_p \tag{1}$$

where M^- is the one-sided maximal operator, and where f^* is the decreasing rearrangement of f. This says that all the powers of $M^- f$ are bounded by those corresponding powers of the decreasing rearrangement of f we call f^*. A related condition is

$$|\{x : M^- f(x) > \lambda\}| \leq |\{x : M^- f^*(x) > \lambda\}|. \tag{2}$$

This expression was proven true by Hardy and Littlewood in the case of one-sided maximal operators (Duoandikoetxea) for functions on compact support. It of course proves the above dialogue-based conjecture for this case.

3.1. General Properties

Maximal operators and decreasing rearrangements are given top billing in our text book of record by Duoandikoetxia, well, almost top billing. They occur in Chapter 2 and the author seems to be saying that you should know these fine operators including decreasing rearrangements even before Hilbert Transforms, Singular Integrals, BMO, weighted inequalities and the T1 theorem.

An interesting property of the Maximal Operator M_d we use for dyadics is that, in a sense we will explore shortly, it is equivalent to other maximal operators. For all of these the decreasing rearrangement is a convenient way to express results in an easily comprehensible way.

3.1.1. s-,d-,g- relates to

A number of equivalence relationships are of the form: A relates to B and B relates to A. This motivates the following definitions of s-relates, d-relates, and g-relates and s-equivalent, d-equivalent g-equivalent. Here s indicates scaling, d indicates dialation, and g indicates graph scaling or more precisely:

Definitions of$A\acute{\succ}B$, $A\bar{\succ}B$,$A\ddot{\succ}B$: We say for two positive functions $f(x), g(x) : \mathcal{R} \mapsto \mathcal{R}$ that f $s-$relates to g or f relates to g by scaling if

$$f\acute{\succ}g \Rightarrow \exists a > 0 : \; af > g; \text{ so } f\acute{\succ}g, \; g\acute{\succ}f \Rightarrow f\acute{\sim}g. \tag{3}$$

f $d-$relates to g or f relates to g by dialation

$$f\bar{\succ}g \Rightarrow \exists a > 0 : \; f(ax) > g(x)\forall x; \text{ so } f\bar{\succ}g, \; g\bar{\succ}f \Rightarrow f\bar{\sim}g. \tag{4}$$

f $g-$relates to g or f relates to g by graph scaling (note different usage of g)

$$f\ddot{\succ}g \Rightarrow f\acute{\succ}g, \text{ and } f\bar{\succ}g; \text{ so } f\ddot{\succ}g, \; g\ddot{\succ}f \Rightarrow f\ddot{\sim}g. \tag{5}$$

So $f\ddot{\succ}g$ can be expressed as follows: there $\exists$ positive numbers $a, b \in \mathcal{R}$ such that for all x

$$af(bx) \geq g(x) \tag{6}$$

Note that the $>$ can be replaced by $\geq$ corresponding to a different choice of a. Also if $b = 1$ holds it follows that $f\acute{\succ}g$ and if $a = 1$ it follows that $f\bar{\succ}g$. We say graph scaling is less restrictive (or perhaps weaker) than either scaling or dialation because functions equalent modulo scaling are automatically equalent modulo graph scaling.

$$\begin{aligned} f \text{ relates to } g \text{ by scaling} &\Rightarrow f \text{ relates to } g \text{ by graph scaling} \\ f \text{ relates to } g \text{ by dialation} &\Rightarrow f \text{ relates to } g \text{ by graph scaling} \end{aligned} \tag{7}$$

Or we might say graph scaling is extensive of scaling and of dialation and the adjective graph is extensive just as the adjective *universal* is extensive as in the common usage *universal health care.*

Before we apply graph scaling to operators, we note the simple result from measure theory with the Borel sigma algebra $\mathcal{B}^n$, $n \in \mathcal{Z}_{>0}$ measurable by the Lebesque-Borel measure indicated by $|.|$ where $a, b > 0$:

$$|\{x \in R^n : bf(ax) > \lambda\}| = \frac{1}{a^n}|\{x \in R^n : f(x) > \frac{\lambda}{b}\}|. \tag{8}$$

In this notation, Lemma 2.12 of Duoandikoetxea says using distribution function

$$\alpha_{f(x)}(\lambda) := |\{x \in \mathcal{R}^n : f(x) > \lambda\}| \tag{9}$$

and

$$|\{x \in \mathcal{R}^n : M'f(x) > \lambda\}| \leq 2^n |\{x \in \mathcal{R}^n : M_d f(x) > \lambda\}| \tag{10}$$

can be written

$$\alpha_{M'f}(\lambda) \leq 2^n \alpha_{M_d f}(4^{-n}\lambda), \text{ or } \alpha_{M_d f} \mathrel{\ddot{\succ}} \alpha_{M'f} \tag{11}$$

so also using the greater generality of M' compared to M_d by definition

$$\alpha_{M'f} \geq \alpha_{M_d f} \mathrel{\ddot{\succ}} \alpha_{M'f} \text{ or } \alpha_{M'f} \mathrel{\ddot{\sim}} \alpha_{M_d f} \text{ or as operator } \alpha_{\mathrm{M'.}} \mathrel{\ddot{\sim}} \alpha_{\mathrm{M_d.}}, \tag{12}$$

on the subspace where these maximal operators exist. Also we also utilize the incompletely proven remarks in Section 2.5 that imply $M \mathrel{\acute{\sim}} M' \mathrel{\acute{\sim}} M''$ (by scaling). Now $A \mathrel{\acute{\sim}} B$ implies $\alpha_A \mathrel{\acute{\sim}} \alpha_B$ implies $\alpha_A \mathrel{\ddot{\sim}} \alpha_B$. so

$$\alpha_{M'.} \mathrel{\ddot{\sim}} \alpha_{M_d.} \mathrel{\ddot{\sim}} \alpha_{M.} \mathrel{\ddot{\sim}} \alpha_{M''.} \tag{13}$$

3.1.2. Decreasing Rearrangements

Definition:$f^*(x) := f(\sigma x)$ such that f is essentially non-increasing and $\mathrm{supp}(f) \subset [0, \infty)$. Note that the non-increasing requirement requires that $\exists \epsilon > 0 : [0, \epsilon) \subset \mathrm{supp}(f)$. Without the ϵ requirement the function would start at 0 and either stay 0 or increase.

Now the distribution is defined as

$$\alpha_f(\lambda) := \{x \in X : f(x) > \lambda\} \tag{14}$$

Lemma: Proof by picture or since f is non-increasing

$$f^*(x) = \inf_{\lambda} \{\alpha_f(\lambda) > x\}. \tag{15}$$

Hardy and Littlewood define decreasing rearrangements by a tasteful

$$f^*(\alpha_f(\lambda)) = \lambda. \tag{16}$$

Exercise: When f^* is strictly decreasing prove the H-L inverse relationship of f^* with α_f

Exercise: show $\alpha_A \mathrel{\acute{\succ}} \alpha_B \Rightarrow A^* \mathrel{\bar{\succ}} B^*$. Show $\alpha_A \mathrel{\bar{\succ}} \alpha_B \Rightarrow A^* \mathrel{\acute{\succ}} B^*$. Show in operator language $M'.^* \mathrel{\ddot{\sim}} M_d.^* \mathrel{\acute{\sim}} M.^* \mathrel{\acute{\sim}} M''.^*$.

4. Comparisons

4.1. Obtaining $M^{-}f$, the Hardy-Littlewood One-Sided Maximal from f

We can let $< f >$ be defined as the average of f over its support

$$\begin{aligned} M^{-}f(x) &= \sup_a < f(.)\chi[a,x)(.) > \\ M^{+}f(x) &= \sup_b < f(.)\chi[x,b)(.) > \end{aligned} \tag{17}$$

Then with the pariety operator P an involution ($P^2 = 1$) with $Pf(x) = f(-x)$ and $P\chi[a,b) = \chi(-b,-a]$,

$$\begin{aligned} PM^{+}f(x) = M^{+}f(-x) &= \sup_b \frac{1}{b+x}\int_{-x}^{b} f(y)dy \\ &= \sup_b \frac{1}{b+x}\int_{-b}^{x} f(-z)dz \\ &= \sup_{-b} \frac{1}{x-b}\int_{b}^{x} f(-z)dz \\ &= M^{-}(Pf)(x). \end{aligned} \tag{18}$$

Obtaining $M^{+} = PM^{-}P$.

4.2. The Hardy-Littlewood H_L Functional at $\Phi(x) = x^p$ and it's Extension.

The following is a partial result using the definition

$$H_L(f(x)) := ||(M^{-}f)(x)||_p^p,\ \ f \in L_1(\mathcal{R}, \mathcal{B}^1, |.|). \tag{19}$$

Recalling:

$$\begin{aligned} M^{-}f(x) &= \sup_a < f(.)\chi[a,x)(.) > \\ M^{+}f(x) &= \sup_b < f(.)\chi[x,b)(.) > \end{aligned} \tag{20}$$

where the average is over the (.) variable.

We wish to make the connection to the uncentered maximal operator

$$M'' = \sup_{a,b} < f(.)\chi[a,b)(x)\chi[a,b)(.) > \tag{21}$$

defined by Duoandikoetxea.

Fix x. Let us use m'', m^+, m^- to correspond to M'', M^+, M^- but without the supremium over a or b thereby resulting in an additional argument or two. So we have

$$(m''f)(x, a, b) = \frac{bm^+f(x, b) + am^-f(x, a)}{b - a} \tag{22}$$

Now consider a sequence (a_m, b_m) such that $m''(x, a_m, b_m)$ is non-decreasing and define

$$\liminf_{m\to\infty}(a_m, b_m) := (\alpha, \beta). \tag{23}$$

Because $f \in L_1$ and $f \to 0$ at $-\infty$ m'' obtains it's extremal value at a finite pair of reals and we have

$$M''f = \frac{\beta m^+f(x, \beta) + am^-f(x, \alpha)}{\beta - \alpha} \tag{24}$$

Letting $\beta/(\beta - \alpha) = w^+(x)$ we have

$$M''f = w^+(x)m^+f(x, \beta) + [1 - w^+(x)]m^-f(x, \alpha) \le w^+(x)M^+ + [1 - w^+(x)]M^-f. \tag{25}$$

A similar argument leads to a choice for b_n and for a_m that cause respectively m^+ and m^- to be independently non-decreasing.

$$(m''f)(x, a_m, b_n) = \frac{b_n m^+f(x, b_n) + a_m m^-f(x, a_m)}{b_n - a_m} \tag{26}$$

Again two bounded limits are obtained in the lim inf at new α, β and

$$M''f = w^-(x)m^+f(x, \beta) + [1 - w^-(x)]m^-f(x, \alpha) \ge w^-(x)M^+f + [1 - w^-(x)]M^-f. \tag{27}$$

Yielding in summary

$$w^+(x)M^+f + [1 - w^+(x)]M^-f \ \ge\ M''f \ \ge\ w^-(x)M^+f + [1 - w^-(x)]M^-f. \tag{28}$$

We can say a little about these weights by omitting the middle term:

$$w^+(x)M^+f + [1 - w^+(x)]M^-f \ \ge\ w^-(x)M^+f + [1 - w^-(x)]M^-f. \tag{29}$$

or

$$(w^+ - w_-)M^+f + (w^- - w^+)M^-f \ \ge\ 0. \tag{30}$$

or

$$(w^+ - w_-)(M^+f - M^-f) \ \ge\ 0. \tag{31}$$

so the weights are in the same order as are M^+f, M^-f.

We can extend $H_L(f)$ with the bounds on $M''$of the form $wM^- + (1-w)PM^-P$, the triangle inequality and with the extension defined to be

$$H_E(f(x)) := ||(M''f)(x)||_p^p \tag{32}$$

$$w^-||(M^-f)||_p^p + (1-w^-)||(PM^-Pf)||_p^p \;\leq\; H_E(f) \leq w^+||(M^-f)||_p^p + (1-w^+)||(PM^-Pf)||_p^p \tag{33}$$

Using $||Pf||_p = ||f||_p$, we obtain with $0 \leq w^+, w^- \leq 1$.

$$w^-||(M^-f)||_p^p + (1-w^-)||(M^-Pf)||_p^p \;\leq\; H_E(f) \leq w^+||(M^-f)||_p^p + (1-w^+)||(M^-Pf)||_p^p. \tag{34}$$

4.3. Rearrangements $f(\sigma x)$

Here we can think of the dyadic description of $f(x)$ and for each family $\mathcal{Q}_k$, with $Q_i, Q_j \in \mathcal{Q}_k$ and its children there is a generator for the group of isometries (ij) that transposes Q_i with Q_j. Here, because we think it is clear, the convention is that as far as properties with which we are concerned, the parent is oblivious to the order of its children i.e. the children are a set. Also if a parent moves, so does its children and subsequent generations. The generators are all those that switch the two children of some parent. Inter parental moves, must occur at a sufficiently higher level that they are one or more intra parental moves. For the case of simple functions defined on a dyadic structure with 2^k bins in the most refined family, the generators are interchange of the children of a common parent for all parents. This generator set permits interchange at the lowest level, as may be verified by the ability of this generator set to sort unique integers assigned to each leaf of the tree in ascending or descending order. We take σ to be one such word made up of letters each consisting of a generator. What we have described can be described more elegantly in terms of Haar bases, but that is beyond the scope of this presentation.

4.4. L^p Norm Invariance Under Decreasing Rearrangements f^*

Norm. $||f^*||_p = ||f||_p$ (because σ preserves measure).

4.5. Theorem(HL) Coincidence of $H_L f^* =$ and $\sup_\sigma H_L(f(\sigma x)), 0 < p < \infty$

$$\begin{aligned} H_L(f(x)) = ||(M^- f)(x)||_p^p = ||(M^- f)(\sigma x)||_p^p \\ H_L(f(\sigma x)) = ||(M^- f)(\sigma x)||_p^p \end{aligned} \tag{35}$$

Theorem II by HL shows that f^* in the discrete case maximizes the left hand side

$$\begin{aligned} H_L(f^*(x)) = \sup_\sigma ||(M^- f)(\sigma x)||_p^p \\ H_L(f^*(x)) = \sup_\sigma H_L(f(\sigma x)) \\ H_L(f^*(x)) = ||(M^- f^*)(x)||_p^p = ||(M f^*)(x)||_p^p \end{aligned} \tag{36}$$

The last observation is also by Duoandkioetxia that for a decreasing function there is no difference between the one-sided operator and M.

Then with results by Riesz,Herz,Bennet and Sharpley

$$c_n (Mf)^*(x) \leq (Mf^*)(x) \leq C_n (Mf)^*(x) \tag{37}$$

And taking norm

$$c_n ||(Mf)^*(x)||_p \leq ||(Mf^*)(x)||_p \leq C_n ||(Mf)^*(x)||_p \tag{38}$$

Then substitution for the middle term

$$c_n ||(Mf)^*(x)||_p \leq ||(M^- f^*)(x)||_p \leq C_n ||(Mf)^*(x)||_p \tag{39}$$

But by (2.5) Duo,

$$c_n ||M' f(x)||_p \leq ||Mf(x)||_p \leq C_n ||M' f(x)||_p \tag{40}$$

And by Lemma 2.12 (Note the upper limit)

$$|\{M' f(x) > \lambda\}| \leq 2^n |\{M_d f(x) > 4^{-n}\lambda\}| \leq 2^n |\{M' f(x) > 4^{-n}\lambda\}| \tag{41}$$

Coincidence and Uniqueness: $H_L(f^*(x)) = H_L(f^*(\sigma x)) \Rightarrow \sigma = 1$ requires proof of injectivity of $M^- f(x)(p)$ and f^* to be strictly decreasing.

Exercise: translate these results using $s-, d-, g-$ relates or equivalence.

4.6. Moreover $\sup_\sigma H_L f(\sigma(x)) = L_p^p[M_- f], p \in Z > 0$

The moral of the story is that except for the one-sided operators, the maximal operators we have encountered M, M', M'', M_d are equal modulo graph scaling including the dyadic operator when operating on f^*. We have shown that the Hardy Littlewood functional that achieves its maximum real value for decreasing rearrangements also our result shows that to within scaling M'' is nearly optimum as is true to within graph scaling the other three maximal operators. If the properties with which you are working are invariant with respect to automorphisms of the real line as are the Banach spaces L_p then modulo graph scaling, you can work with the decreasing rearrangement.

We can say more, but it would be nice to have more results.

Remark Dyadic approach to decreasing rearrangements: does not always optimize on deceasing rearrangement, based on above remarks they are close in that sence.

Exercise: Find four numbers illustrating this point.

4.7. More

More results in the future are expected.

5. Acknowledgements

Thanks are due to Prof. Marie Christina Pereya for suggesting a project for this semester course and for the work she has invested in preparing motivating lectures and for her encouragement. For this paper itself, Professor Pereya had pointed out in earlier drafts a number of serious errors or points that needed proof or clarification which hopefully have been corrected and/or provided here. The approach taken in this project has been self discovery, and when such discovery takes place when results are already known or felt by at least a few and/or already published or hinted at I make no claim of originality nor have I tried to prove originality. Thanks also to Justin Pati for pointing out the computer science construction of permutations is also a barycentric decomposition. Both Pati and Pereya also noted a Haar basis connection with this work.