THE VARIATION OF THE CENTERED MAXIMAL FUNCTION

JOÃO P. G. RAMOS (AFTER O. KURKA)

Abstract. In these notes, we explore a bit the ideas of O. Kurka [1] of how to prove that, given a $f \in BV[\mathbb{R})$, then $Mf \in BV(\mathbb{R})$ with $BV-$norm bounded by a constant multiple of that of f.

1. Introduction and Main Ideas

First of all, it is useful to answer *why* the dramatic difference between the centered and uncentered cases.

The straight-out answer to this question is fairly simple: as presented by Marco, the uncentered maximal function satisfies a "magic" property: if it attains a (local) maximum at a point, then so does f, and both of them coincide at that point.

This fantastic property unfortunately does not hold for the centered maximal function: think, for instance, about what happens when we take $f = \chi_{(0,1)} + \chi_{(2,3)}$. Nonetheless, as Kurka [1] showed, the property *still* holds, but now with a constant much worse than in the uncentered case: in the latter, $C_{uncentered} = 1$, while in the former one one only knows that the best constant satisfies $C_{centered} \leq 240,004$.

If we cannot use Aldaz and Pérez-Lázaro's ideas, then what should we do to prove it? First of all, one does not even need to move much further away from their paper to get a flavour of what is being done here: besides proving the mapping from BV to itself of the uncentered maximal function, they prove also that the uncentered maximal function actually *improves* the regularity of a function: if $f \in BV$, then its uncentered maximal function is an *absolutely continuous* function.

The way they do it is by contradiction, usind the Banach-Zarecki Theorem (as pointed in the end of last lecture by Marco): suppose $m(N) = 0$ but $m^*(M_{uncentered}f(N)) > 0$. Then, for a suitably fine partition of the real line $\mathcal{P}$, one might refine it to obtain

$$\mathcal{V}(f;\mathcal{P}) + \frac{1}{8}m^*(M_{uncentered}f(N)) \leq \mathcal{V}(f;\mathcal{P}'),$$

where $\mathcal{P}'$ is the refinement of $\mathcal{P}$. It is somehow difficult to realise, but a similar procedure is being used in Kurka's work: roughly, his techniques might be divided into the following steps:

(A) For a given partition, one would like to decompose it into *peak* structures. One might further distinguish them into essential and non-essential peaks. As the name suggests, the non-essential ones may be handled in an easy way, leaving us to work with the essential peaks.
(B) As these essential peaks might be randomly distributed within the partition under consideration, one would like then to divide them into subsets, so that they look fairly similar inside these subsets. One is then able to prove that the variation of the centered maximal function within each of those sets is, in a certain sense, *almost* controlled by the variation of the function within those sets.
(C) The division into those sets is done by a layer dyadic decomposition according to the radius of the top of a peak. As the decomposition is dyadic, it makes sense to talk about the *scale* of these peaks.
(D) The next step is then not more than Aldaz and Pérez-Lázaro's idea for the $BV \to AC$ mapping of the uncentered, together with an induction on scales argument: roughly, if

there is a partition $\mathcal{P}$ of the real line whose elements are all separated by at least L, then one might refine it so that the new partition $\mathcal{P}'$ has elements at least $\frac{L}{1024}$ appart from one another, and

$$\mathcal{V}(f;\mathcal{P}) + \frac{1}{5}V_n \leq \mathcal{V}(f;\mathcal{P}'),$$

where V_n denotes the variation of the centered maximal function over the peaks of scale n.

Of course, there are many inaccuracies and oversimplifications in the prototypical process described above, and things get much more technical once one starts to work hands-on to understand Kurka's work, but the guideline for possible generalizations should be along these lines. Moreover, the author remarks in his work that this scheme did not arise from nothing, from the beginning: the original idea was to use the first Lemma below to prove it, possibly even with the (conjectured) sharp constant. His attemps, however, were not fruitful, and the procedure above was the way he found to do it in the end.

The rest of the notes is divided as follows: In Section 2, we define rigorously our basic peak structures and how to divide them into simpler groups, together with the specific estimates one has for those groups. In Section 3 we state our main strategy to deal with the interaction between different scales, and in Section 4 we give a sketch of proof (in one of the cases) of the induction on scales Lemma we will be using.

2. Peaks and groups of peaks

First of all, let us focus on what we want: we wish to bound the variation of the centered maximal function – denoted from now on by M – over a given partition $\mathcal{P}$ by a constant times the variation of the function under consideration. Let this partition be $\mathcal{P} = \{x_1 < x_2 < x_3 < \cdots x_N\}$. We also assume that this partition is not redundant, in the sense that $Mf(x_{2i+1}) - Mf(x_{2i})$ has always the same sign as $Mf(x_{2i}) - Mf(x_{2i-1})$.

We also remark that, without loss of generality, one has that $Mf(x_1) \leq Mf(x_2) \geq Mf(x_3) \leq \cdots$. This leads directly to the definition of a peak:

Definition 1. *A system of 3 points* $\mathbf{p} = \{p < r < q\}$ *is a* peak *if* $Mf(r) > \max\{Mf(p), Mf(q)\}$. *Moreover, we call a peak* $\mathbf{p}$ essential *if*

$$\sup_{y\in[p,q]} f(y) \leq Mf(r) - \frac{1}{4}\mathcal{V}(\mathbf{p}).$$

Here, $\mathcal{V}(\mathbf{p}) = 2Mf(r) - Mf(q) - Mf(p)$.

It is an easy task – that we leave to the reader – to verify that, if $\mathbf{P}$ is an arbitrary collection of non-essential peaks, then $\mathcal{V}(\mathbf{P}) \leq 8\mathcal{V}(f)$. So we will focus only on the essential ones. It is also easy to notice that, for an essential peak $\mathbf{p}$, the minimal radius $w(r)$ such that $⨍_{r-w(r)}^{r+w(r)} f = Mf(r)$ is well-defined, and $r - w(r) < p < q < r + w(r)$.

We can now define the decomposition of our peaks: suppose we reduced our partition $\mathcal{P}$ to one only with essential peak systems, which we call again $\mathcal{P}$. Denote

$$50L_0 = \max\{w(r_i), r_i \text{ is the top of } \mathbf{p}_i \subset \mathcal{P}\}.$$

We perfom the promised dyadic decomposition: $L_n = 2^{-n}L_0$, and we first let $\mathbf{P}^n$ be the set of peaks $\mathbf{p}_i$ for which $w(r_i) \in (25L_n, 50L_n]$. Still, these might be quite randomly distributed across the real line, and we therefore further specify them into the sets $\mathbf{P}^n_k \subset \mathbf{P}^n$ such that $r_i \in (kL_n, (k+1)L_n]$. This might be understood as a reduction to a suitable dyadic model.

To prove that it is at least interesting giving this decomposition a chance, we have the following:

Lemma 1. *Let* $[a, b]$ *be an interval of length* L *and* $\mathbf{P} = \{\mathbf{p}_i,\ 1 \leq i \leq m\}$ *a system of essential peaks such that:*

- $r_i \in [a,b], i = 1,\dots,m$;
- $q_i \leq p_{i+1}, i = 1,\dots,m-1$;
- $w(r_i) \in (25L, 50L], i = 1,\dots,m$.

Then there are $s < u < v < t$ such that

- $a - 50L < s < t < b + 50L$;
- $u - s, t - v \geq 4L, u - v \geq L$

and also

$$\min\{f(s), f(t)\} - \fint_u^v f \geq \frac{1}{12}\mathcal{V}(\mathbf{P}).$$

Although the proof of this Lemma is interesting, we cannot restrict our attention to it here. We remark two things, however: the first one is that, after this Lemma, we basically never use any property of the maximal function again. That is, the next steps of the proof are completely general and do not depend on the particular structure of the maximal function. The second is that the proof of this fact uses essentially only one property of the maximal function, which we bring to light due to its possible importance in future related work:

Property 1. *Let $r \in \mathbb{R}$ and $w(r)$ the radius of $Mf(r)$, as defined above. Let moreover $r - w(r) < p < r$ be such that $Mf(p) \leq Mf(r)$. Then we can find $2p - (r - w(r)) < t < r + w$ such that*

$$f(t) \geq Mf(p) + \frac{Mf(r) - Mf(p)}{r - p} \cdot w(r) \geq Mf(r).$$

A similar statement holds for if $r < s < r + w(r)$ and $Mf(q) \leq Mf(r)$.

By using Lemma 1 with the sets $\mathbf{P}_k^n$, we can already get the estimate

$$\mathcal{V}(\mathbf{P}_k^n) \leq 12\mathcal{V}(f; ((k-50)L_n, (k+51)L_n]).$$

Unfortunately, this is not yet good enough in order to sum up and obtain the result: in fact, this estimate only takes care of single-scale terms, and ignores the interaction between scales. This leads to the idea of trying to bootstrap the scale $n+1$ once one has bounds for scale n.

3. Scale interaction and dichotomy

As promised, we will now get rid completely of the maximal function:

Lemma 2. *Let Λ_k^n be a doubly-indexed sequence of nonnegative real numbers, of which only a finite set is non-zero. Let also $L_0 > 0$ and $L_n = 2^{-n}L_0$. Suppose now that, whenever $\Lambda_k^n > 0$, then there are reals $s < u < v < t$ such that*

- $(k-50)L_n \leq s < t \leq (k+51)L_n$;
- $u - s, t - v \geq 4L_n, u - v \geq L_n$,

and, in addition,

$$\min\{f(s), f(t)\} - \fint_u^v f \geq \Lambda_k^n.$$

Then

$$\sum_{n,k} \Lambda_k^n \leq 20000\mathcal{V}(f).$$

Obviously, this is *exactly* what we need to proceed. To prove this, we have, however, to distinguish two cases: the first models when $f(s), f(t)$ are particularly big inside the interval $((k-50)L_n, (k+51)L_n]$, and the other when the average $\fint_u^v f$ is particularly small in comparison with the rest of the points in the interval. Rigorously stated, we have the following dichotomy:

Lemma 3. *Let $n \geq 0, k \in \mathbb{Z}, \Lambda_k^n > 0$. Then one of the following holds:*

(A) *There is a 6-point system $(k-50)L_n \leq s < \alpha < \beta < \gamma < \delta < t \leq (k+51)L_n$ such that they are all at least L_n apart from one another, $\gamma - \beta = 2L_n$, and*

$$\min\{f(s), f(t)\} - \max\left\{\fint_\alpha^\beta f, \fint_u^v f\right\} \geq \frac{1}{2}\Lambda_k^n.$$

(B) *There is a 6-point system* $(k-50)L_n \leq \alpha < \beta < u < v < \gamma < \delta \leq (k+51)L_n$ *such that they are all at least* L_n *apart from one another and*

$$\min\left\{\fint_\alpha^\beta f, \fint_\gamma^\delta f\right\} - \fint_u^v f \geq \frac{1}{2}\Lambda_k^n.$$

Therefore, the most reasonable thing to do is to call the pairs (n,k) such that condition (A) holds as belonging to the set $\mathcal{A}$, and the same for $\mathcal{B}$. Furthermore, as we want to use some disjointness within scales, we further divide the k in a same scale n into the sets $\mathcal{A}_K^n = \{k; (n,k) \in \mathcal{A} \text{ and } k = K \mod 200\}$.

We remark that this division is a merely technical asumption, as a thorough scrutiny of the proof contained in [1] reveals that both cases share the same basic outline for the proof of their boundedness. However, the approaches seem not to be treatable at the same time by the same technique, but this will not prevent us from only visiting the proof of boundedness for systems in $\mathcal{A}$.

4. An Induction on scales argument

The strategy now is simple: in order to run the argument and take advantage of some sort of disjointness, we will induct on η, where $n = 10\eta + b$, and $k = K \mod 200$. Those assumptions being made, we move on to the following procedure: suppose that we have a $1024L_n-$separated system of points

$$X_1 < U_1 < V_1 < X_2 < \cdots < X_M.$$

Then we claim that we can find a system of L_n-separated points

$$x_1 < u_1 < v_1 < x_2 < \cdots < x_m$$

such that

$$\sum_{i=1}^{m-1}\left[f(x_i) + f(x_{i+1}) - 2\fint_{u_i}^{v_i} f\right] \geq \sum_{I=1}^{M-1}\left[f(X_I) + f(X_{I+1}) - 2\fint_{U_I}^{V_I} f\right] + \frac{1}{5}\sum_{k\in\mathcal{A}_K^n}\Lambda_k^n. \tag{1}$$

This is more or less what we referred to in the introduction as point (D) in our strategy. Of course, this procedure is merely a device to induct: we will generally assume that the system $X_1 < U_1 < \cdots < X_M$ already satisfies an inequality like

$$\sum_{I=1}^{M-1}\left[f(X_I) + f(X_{I+1}) - 2\fint_{U_I}^{V_I} f\right] \geq \frac{1}{5}\sum_{\substack{l\leq n-10,\\ k\in\mathcal{A}_K^l}}\Lambda_k^l.$$

Of course, as we are dealing here with finite structures, there is an N_0 such that $\mathcal{A}_K^n \neq \emptyset \Rightarrow n \leq N_0$. We use then our procedure until we reach this N_0. What we obtain, in the end, is that

$$\sum_{\substack{n=N \mod 10,\\ k=K \mod 200}}\Lambda_k^n \leq 5\mathcal{V}(f).$$

Summing in N over all residue classes $\mod 10$ and in K over all residue classes $\mod 200$, one obtains the desired result, with constant equal to 10000 (the other 10000 coming from the case we are omitting).

Proof of the induction argument in (1). Step 1: For every $k \in \mathcal{A}_K^n$, one selects a L_n-separated system $(k-50)L_n \leq s_k < \alpha_k < \beta_k < t_k \leq (k+1)L_n$, such that

$$\min\{f(s_k), f(t_k)\} - \fint_{\alpha_k}^{\beta_k} f \geq \frac{1}{2}\Lambda_k^n$$

and, for all $1 \leq I \leq M$,

$$\operatorname{dist}(X_I, (\alpha_k, \beta_k)) \geq L_n.$$

This is possible due to the separation conditions we have on our systems. Finally, call an interval (a,b) *orthogonal* to some $k \in \mathbb{Z}$ ($(a,b) \perp k$) if (a,b) is at least L_n apart from $((k-50)L_n, (k+51)L_n]$.

Step 2: We will skip the details, but, even without using the systems $s_k < \alpha_k < \beta_k < t_k$ constructed above, one is able to prove the following: there is a subset $\mathcal{S} \subset \mathcal{A}_K^n$ for which one can find a L_n-separated system $y_1 < c_1 < d_1 < \cdots < y_j$ such that

- $(c_i, d_i) \perp l$, $\forall l \in \mathcal{A}_K^n \backslash \mathcal{S}$, $\forall i = 1, ..., j-1$;
- $\operatorname{dist}(y_i, (\alpha_l, \beta_l)) \geq L_n$, $\forall l \in \mathcal{A}_K^n \backslash \mathcal{S}$, $\forall i = 1, ..., j$,

and

$$\sum_{i=1}^{j-1} \left[f(y_i) + f(y_{i+1}) - 2 \fint_{c_i}^{d_i} f \right] \geq \sum_{I=1}^{M-1} \left[f(X_I) + f(X_{I+1}) - 2 \fint_{U_I}^{V_I} f \right] + \frac{1}{5} \sum_{k \in \mathcal{S}} \Lambda_k^n.$$

Step 3: Finally, one adds the remaining elements one by one: let $y_1 < c_1 < d_1 < \cdots < y_j$ obtained from the previous step, and pick $k \in \mathcal{A}_K^n \backslash \mathcal{S}$. We know, by Step 2, that things are well spaced, and that is, essentially, what will allow us to build our new scale.

In fact, as $(c_i, d_i) \perp k \Rightarrow$ the interval $((k-50)L_n, (k+51)L_n)$ is contained in $\mathbb{R} \backslash \cup_i [c_i, d_i]$. Let then $y_{t(k)}$ be the element belonging to the same connected component that covers $((k-50)L_n, (k+51)L_n)$. Now, replace $y_{t(k)}$ by a system $s_k' < \alpha_k < \beta_k < t_k'$, where α_k, β_k are the ones selected above. As one would expect, s_k' equals s_k when either $y_{t(k)} > \alpha_k - L_n$ or $f(y_{t(k)}) < f(s_k)$, and equals $y_{t(k)}$ otherwise. One picks t_k' in a completely analogous manner.

Step 4: Checking that the system obtained by this substitution has at least the variation of the previous one $+\Lambda_k^n$ is a simple task, and we therefore skip it. One has also, by the way we picked our points, that this newly obtained system satisfies the properties of Step 2. We then run Step 3 over, and this yields a system with the required properties, for all $k \in \mathcal{A}_K^n$. This finishes the proof of the induction scheme, and, therefore, of the main Theorem. □

References

[1] O. Kurka, *On the variation of the Hardy-Littlewood maximal function,* Annales Academiae Scientiarum Fennicae Mathematica **40** (2015), 109–133.