The Journal of Fourier Analysis and Applications
Volume 2, Number 5, 1996

A Search for Best Constants in the Hardy–Littlewood Maximal Theorem

Ron Dror, Suman Ganguli, and Robert S. Strichartz

ABSTRACT. Let $Mf(x) = \sup(1/2r)\int_{x-r}^{x+r}|f(t)|\,dt$ *be the centered maximal operator on the line. Through a numerical search procedure, we have conjectural best constants for the weak-type 1–1 estimate* (3/2) *and the* L^p *estimate (the constant* $B(p,1)$ *such that* $M(|x|^{-1/p}) = B(p,1)|x|^{-1/p}$*). We prove that these constants are lower bounds for the best constants and discuss the numerical evidence for the conjectures.*

1. Introduction

L^p estimates for specific operators are statements with empirical content, so it should be possible to obtain information about them, at least conjectural, through numerical experiments. As far as we know, this approach has not been tried before. This paper reports on a "pilot study" of the one-dimensional centered maximal operator. Since the L^p mapping properties are given by the well-known theorem of Hardy and Littlewood, we concentrated on the problem of finding the best constants in the weak-type $(1, 1)$ estimate and the L^p estimates $(1 < p < \infty)$. A more challenging project would be to study an operator for which the L^p mapping properties are not known.

The maximal operator is a positive operator, so it suffices to consider only nonnegative functions. Also, because of positivity, it is not necessary to do computations with extreme accuracy to obtain reliable experimental results. It would seem prudent, at first, to limit experimental studies to positive operators, for these reasons. On the other hand, the maximal operator is nonlinear. It is possible that linear operators might be easier targets for the experimental approach.

As a result of our experiments, we obtained conjectures for the best constants. In §§2 and 3 we present proofs that these conjectural constants are lower bounds for the best constants. These sections are presented in traditional mathematical form, with no reference to the experiments. It is quite possible that these results could have been discovered without using the experimental approach, but in fact they emerged directly from an analysis of the experimental results. (After this work was completed, the paper by Christ and Grafakos [1] appeared, which also contains a proof of our Theorem 3.2.) To this extent, the experimental approach has already been successful. Of course, we will not be able to claim an impressive victory until someone succeeds in proving that our conjectured bounds are optimal.

Note Added June 1996. A recent preprint of Grafakos, Montgomery–Smith, and Motrunich [4] proves our conjectured L^p bound for a class of "bell-shaped" functions.

Math Subject Classification. Primary 42B25.
Keywords and Phrases. Hardy–Littlewood Maximal Theorem, best constants.
Acknowledgements and Notes. The first and second authors' research was supported by the National Science Foundation through the REU Program. The third author's research was supported by the National Science Foundation Grant DMS–9303718.

ISSN 1069-5869

One disappointment we have to report is that we are not able to say anything about the higher dimensional case. The complexities that arise in computing the maximal operator even in two dimensions make it seem unlikely that a direct search could yield anything useful. It was our hope that the one-dimensional results would shed some light on the situation in higher dimensions. They do not. We attempted to mimic the one-dimensional construction for the weak-type $(1, 1)$ estimate, but the results were too inconclusive to report. For the L^p bound, there is an obvious generalization to $\mathbb{R}^n$, but we show in §3 that it is inadequate. The existence of L^p bounds that are independent of n was established in Stein [6]. See also Stein and Stromberg [8].

In §§4 and 5 we discuss the computer search that led to the results presented in §§2 and 3 and the extent to which the experimental evidence supports our conjectures. Strictly speaking, there are no mathematical theorems in these sections. Nevertheless, they contain some of the main ideas we wish to communicate.

Notation. We write

$$Mf(x) = \sup_{r>0} \frac{1}{\Omega_n r^n} \int_{|y|\leq r} |f(x+y)|\, dy$$

for the centered maximal operator in $\mathbb{R}^n$, where Ω_n denotes the volume of the unit ball in $\mathbb{R}^n$. We let ω_{n-1} denote the $(n-1)$-dimensional area of the unit sphere in $\mathbb{R}^n$. We let $A(1, n)$ denote the smallest constant in the weak-type $(1, 1)$ estimate

$$|\{x : Mf(x) \geq s\}| \leq A(1, n)\|f\|_1/s$$

and $A(p, n)$ for $1 < p < \infty$ denote the smallest constant in the L^p estimate

$$\|Mf\|_p \leq A(p, n)\|f\|_p.$$

The existence of such bounds can be found in many standard references, such as [2] by de Guzman and [5, 7] by Stein. Related work on the uncentered maximal function is reported by Grafakos and Montgomery-Smith in [3]. □

2. Bounds for the Weak-Type Estimate

It is well known that we can replace the space L^1 by the space $\mathcal{M}$ of finite measures in the weak-type estimate without changing the norm

$$|\{x : M\mu(x) \geq s\}| \leq A(1, n)\|\mu\|/s. \tag{2.1}$$

Without loss of generality we may assume that μ is a probability measure, and by dilation invariance we may assume $s = 1$. Thus the constant $A(1, n)$ is just the sup of $|\{x : M\mu(x) \geq 1\}|$ as μ varies over all probability measures. Also, a fairly routine argument shows that we can restrict attention to finite discrete measures $\mu = \sum_{k=1}^{N} a_k\delta(x - x_k)$ (this is true even though M is not continuous in the weak topology). The reason is roughly as follows. By sacrificing ε we can assume that f is a continuous function of compact support and in the computation of Mf the sup is taken over all r satisfying $r \geq r_0$ for some fixed $r_0 > 0$; then approximate all integrals by Riemann sums.

When $n = 1$, it is known that $A(1, 1) \leq 2$. The same estimate works for the uncentered maximal function, and in that case the best constant is exactly 2. It seems unlikely that the smaller centered maximal function would need the same bound. We conjecture that $A(1, 1) = 3/2$.

Theorem 2.1.
$A(1, 1) \geq 3/2$.

Proof. For any N, consider the probability measure $\mu = \sum_{k=0}^{N-1}(1/N)\delta(x - 3k/2N)$. We claim that $M\mu(x) \geq 1$ exactly on the interval $(-1/2N, (3/2) - (1/N))$ of length $(3/2) - (1/2N)$, which will give the desired estimate as we let $N \to \infty$.

To prove the claim, observe that for any point x in the interval $((3k-1)/2N, (3k+1)/2N)$, we can take $r = |x - 3k/2N|$ and obtain an interval of length $2r \le 1/N$ containing a single point mass of weight $1/N$; hence $M\mu(x) \ge 1$. On the other hand, for a point x in the interval $((3k+1)/2N, (3k+2)/2N)$ (for $k \le N-2$) we can take $r = \max(x - 3k/2N, (3(k+1)/2N) - x)$ and obtain an interval of length $2r \le 2/N$ containing two point masses (at $3k/2N$ and $3(k+1)/2N$) of total weight $2/N$; hence $M\mu(x) \ge 1$. These two types of intervals fill up the entire interval $(-1/2N, (3/2) - (1/N))$ as claimed. □

The two types of intervals generated in the proof are nonoverlapping. Furthermore, there are no intervals where $M\mu(x) \ge 1$ because of a choice of r that engulfs three or more point masses. It is possible to give more complicated examples involving such interactions, but they do not yield a better bound for $A(1, 1)$. This will be discussed further in §4. Notice that we do not produce a single measure that attains the value 3/2 for $|\{x : M\mu(x) \ge 1\}|$, and we conjecture that no such extremal exists. Also, it would be a mistake to pay attention to the fact that the sequence of measures that we give is converging weakly to the function $2/3\chi([0, 3/2])$. For this limit function the set $\{x : Mf(x) \ge 1\}$ is empty. To obtain a sequence of functions rather than measures you should put tall skinny spikes of area $1/N$ around each of the points $3k/2N$.

3. Lower Bounds for L^p Estimates

We begin by stating a result that follows easily by a homogeneity argument.

Lemma 3.1.

Let $f(x) = |x|^{-n/p}$ for $p > 1$. Then there exist constants $B(p, n)$ and r_p (also depending on n) such that

$$Mf = B(p, n)f \tag{3.1}$$

and moreover

$$Mf(x) = \frac{1}{\Omega_n(r_p|x|)^n} \int_{|t-x|\le r_p|x|} f(t)\, dt. \tag{3.2}$$

For large n and p we may have $B(p, n) = 1$ and $r_p = 0$. The following theorem also appears in [1] by Christ and Grafakos.

Theorem 3.2.

For $1 < p < \infty$ we have

$$A(p, n) \ge B(p, n). \tag{3.3}$$

Proof. Fix p and define

$$f_N(x) = |x|^{-n/p}\chi(1 \le |x| \le N). \tag{3.4}$$

Then we have

$$\|f_N\|_p = (\omega_{n-1} \log N)^{1/p}. \tag{3.5}$$

We compute a lower bound for Mf_N in the region $1 \le |x| \le N/(1 + r_p)$ by taking the average over the ball of radius $r_p|x|$ about x. Note that the condition $|x| \le N/(1 + r_p)$ implies that this ball lies entirely in $|x| \le N$. Thus

$$Mf_N(x) \ge \frac{1}{\Omega_n(r_p|x|)^n}\left[\int_{|t-x|\le r_p|x|} |t|^{-n/p}\, dt - \int_{|t|\le 1} |t|^{-n/p}\, dt\right].$$

By the lemma we have

$$Mf_N(x) \ge B(p, n)|x|^{-n/p} - c|x|^{-n} \quad \text{in} \quad 1 \le |x| \le N/(1 + r_p),$$

where $c = p'/nr_p^n$, or in other words,

$$Mf_N(x) \geq B(p,n) f_{N/(1+r_p)}(x) - g(x), \tag{3.6}$$

where $g(x) = c|x|^{-n}\chi(|x| \geq 1)$ is in L^p. By (3.5) we obtain

$$\|Mf_N\|_p \geq B(p,n)\omega_{n-1}^{1/p}\big(\log N - \log(1+r_p)\big)^{1/p} - \|g\|_p;$$

hence,

$$\|Mf_N\|_p / \|f_N\|_p \geq B(p,n)\left(1 - \frac{\log(1+r_p)}{\log N}\right)^{1/p} - \|g\|_p/(\omega_{n-1}\log N)^{1/p}.$$

Thus $\lim_{N\to\infty} \|Mf_N\|_p/\|f_N\|_p \geq B(p,n)$. □

When $n = 1$ we conjecture that $A(p,1) = B(p,1)$. In that case we give an implicit equation for r_p that enables us to understand the behavior of $B(p,1)$.

Lemma 3.3.

Let $n = 1$. Then r_p is the unique solution in $r > 1$ of

$$(1 - r/p)^p(r+1) = (1 + r/p)^p(r-1). \tag{3.7}$$

Proof. For $f(x) = |x|^{-1/p}$ we have

$$B(p,1) = Mf(1) = \sup_r (1/2r)\int_{1-r}^{1+r} |t|^{-1/p}\,dt$$

and r_p is the value of r where the sup is attained. Since

$$\frac{1}{2}\left((1+t)^{-1/p} + (1-t)^{-1/p}\right) > 1 \quad \text{for} \quad 0 < t < 1,$$

it is clear that the sup is attained in the region $r > 1$. Thus we have the calculus problem of maximizing

$$\begin{aligned} g(r) &= (1/2r)\left(\int_0^{1+r} t^{-1/p}\,dt + \int_0^{r-1} t^{-1/p}\,dt\right) \\ &= (p'/2r)\left((r+1)^{1/p'} + (r-1)^{1/p'}\right). \end{aligned} \tag{3.8}$$

Note that $\lim_{r\to\infty} g(r) = 0$, so the maximum is attained. We have

$$g'(r) = (1/2r)\left((r+1)^{-1/p} + (r-1)^{-1/p}\right) - (p'/2r^2)\left((r+1)^{1/p} + (r-1)^{1/p'}\right) \tag{3.9}$$

and $g'(1) = +\infty$, so $r = 1$ is not the maximum. Thus r_p is a solution of $g'(r) = 0$, and after some algebraic manipulations this equation becomes (3.7). It is easy to see that (3.7) has only one solution since the derivative of the right side is greater than the derivative of the left side. □

In particular, $r_2 = 2/\sqrt{3}$ and $g(r_2) = \sqrt[4]{27}/\sqrt{2}$. For p an integer, r_p is a solution of a polynomial equation. In general, we can easily compute numerical approximations to r_p and $g(r_p)$ using Newton's method. The solution of (3.7) requires that $r < p$ since the right side is always positive. In fact, it is easy to compute the limits $r_1 = \lim_{p\to 1} r_p$ and $r_\infty = \lim_{p\to\infty} r_p$. We have $r_1 = 1$, and r_∞ is the solution to

$$e^{-r}(r+1) = e^r(r-1) \tag{3.10}$$

or, equivalently,

$$r = \coth r \qquad (r_\infty \approx 1.1996786). \tag{3.11}$$

For large p we have the asymptotic behavior

$$r_p = r_\infty - \frac{1}{3}\frac{r_\infty}{r_\infty^2 - 1}\frac{1}{p^2} + O\Big(\frac{1}{p^3}\Big), \qquad p \to \infty, \tag{3.12}$$

and for small p we have

$$r_p = 1 + (p-1)/4 + O((p-1)^2), \qquad p \to 1^+. \tag{3.13}$$

These results follow from (3.7) and some routine calculations. Then by substituting these results for r_p in (3.8) we obtain

$$B(p,1) = 1 + \frac{1}{p}\Big(1 - \frac{(r_\infty + 1)\log(r_\infty + 1) + (r_\infty - 1)\log(r_\infty - 1)}{2r_\infty}\Big) + O(1/p^2) \quad \text{as} \quad p \to \infty, \tag{3.14}$$

$$B(p,1) = 1/(p-1) + \frac{1}{2}\log(p-1) + O(1) \quad \text{as} \quad p \to 1. \tag{3.15}$$

We can also see that r_p is increasing and $B(p,1)$ is decreasing as a function of p. It is somewhat easier to do this if we introduce the variable $s = 1/p$. Then (3.7) is equivalent to

$$\log(1 - rs) + s\log(r+1) = \log(1 + rs) + s\log(r-1), \tag{3.16}$$

which shows $rs < 1$. Differentiating (3.16) and simplifying (also using (3.16)) we obtain

$$\frac{dr}{ds} = \frac{-(r^2-1)(1-r^2s^2)}{2(rs)^2(1-s^2)}\Big(\log\frac{1-rs}{1+rs} + \frac{2rs}{1-(rs)^2}\Big), \tag{3.17}$$

which is negative since

$$\log\frac{1-x}{1+x} + \frac{2x}{1-x^2} > 0 \quad \text{for} \quad 0 < x < 1.$$

Thus r_p is increasing. In principle we could use the same method to show $B(p,1)$ is decreasing, but the result is exceedingly complicated. Instead, since $B(p,1) = \sup_{r>1} g(r,s)$ for $g(r,s) = \frac{1}{1-s}\frac{1}{2r}((r+1)^{1-s} + (r-1)^{1-s})$, it suffices to show $\frac{\partial g}{\partial s}(r,s) > 0$ for all (r,s) satisfying $1 \le r \le r_\infty$ and $0 < s < 1$. But this follows easily since $\frac{1}{1-s} - \log(r+1) > 0$ under these restrictions (using $r_\infty < e - 1$).

We discuss briefly the case $n > 1$.

Lemma 3.4.

If $p < n/(n-2)$, then $B(p,n) > 1$.

Proof. Consider the spherical average operator

$$S_r f(x) = \frac{1}{\omega_{n-1}}\int_{S^{n-1}} f(x + ru)\,du. \tag{3.18}$$

The result will follow if we can show $S_r f(x) > f(x)$ for all r satisfying $0 < r < \varepsilon$ for some ε, and $|x| = 1$, where $f(x) = |x|^{-ns}$. But then $S_r f(x)$ is just

$$h(r) = \frac{\displaystyle\int_0^\pi (1 + 2r\cos\theta + r^2)^{-ns/2}(\sin\theta)^{n-2}\,d\theta}{\displaystyle\int_0^\pi (\sin\theta)^{n-2}\,d\theta}. \tag{3.19}$$

We have $h(0) = 1$ and

$$h'(0) = -ns \frac{\displaystyle\int_0^\pi \cos\theta (\sin\theta)^{n-2}\, d\theta}{\displaystyle\int_0^\pi (\sin\theta)^{n-2}\, d\theta} = 0$$

for $n \geq 2$. But

$$h''(0) = ns \frac{\displaystyle\int_0^\pi \left[(ns+2)\cos^2\theta - 1\right](\sin\theta)^{n-2}\, d\theta}{\displaystyle\int_0^\pi (\sin\theta)^{n-2}\, d\theta} = ns(s-(n-2)/n),$$

so $h''(0) > 0$ if $p < n/(n-2)$. This shows $h(r) > 0$ for $0 < r < \varepsilon$ as claimed. □

When $n = 3$ we have

$$h(r) = \frac{(1+r)^\lambda - |1-r|^\lambda}{2\lambda r}$$

for $\lambda = 2 - 3s$, which means $1 \leq \lambda < 2$ in the region $p \geq 3$ where the lemma does not apply. When $p = 3$ we have $h(r) = 1$ for $0 \leq r \leq 1$ and $h(r) < 1$ for $r > 1$, so $B(3,3) = 1$; while for $p > 3$ we have $h(r) < 1$ for all $r > 0$ (this is trivial for $r > 1$ and follows from the inequality $(1+r)^\lambda < 2r\lambda + (1-r)^\lambda$ for $0 < r < 1$ and $1 < \lambda < 2$), so again we have $B(p,3) = 1$, and now $r_p = 0$. It seems likely that $B(p,n) = 1$ whenever $p \geq n/(n-2)$.

In these cases, Theorem 3.2 gives no information, since $A(p,n) > 1$.

4. Computational Aspects of the Search ($p = 1$)

As described in §2, the constant $A(1,n)$ is the sup of $|\{x : M\mu(x) \geq 1\}|$ as μ varies over all finite discrete probability measures. Given such a finite discrete probability measure $\mu = \sum_{k=1}^N a_k \delta(x - x_k)$, let $E = \{x : M\mu(x) \geq 1\}$. We describe our method for computing $|E|$ in the one-dimensional case, but the method generalizes easily to higher values of n. We implemented it on a computer for $n = 1$ and $n = 2$.

When $n = 1$,

$$M\mu(x) = \sup_{r>0} (2r)^{-1} \mu([x-r, x+r]). \tag{4.1}$$

Label the point masses so that $x_1 < x_2 < \cdots < x_N$. Given indices m and n satisfying $1 \leq m \leq n \leq N$, let $a = \sum_{k=m}^n a_k$. If $a \geq x_n - x_m$, the point $(x_m + x_n)/2$ is guaranteed to lie in E, since taking $r = (x_n - x_m)/2$ gives $M\mu((x_n + x_m)/2) = 1$. Moreover, if a is strictly greater than $x_n - x_m$, a certain interval centered at $(x_n - x_m)/2$ is guaranteed to lie in E.

Specifically, let $I_{m,n}$ be the interval $[x_n - a/2, x_m + a/2]$ if $x_n - a/2 \leq x_m + a/2$, and the empty interval otherwise. $I_{m,n}$ lies in E for all $1 \leq m \leq n \leq N$. The maximum in (4.1) clearly occurs when $|x_j - x| = r$ for some j. Thus, $E = \bigcup_{1 \leq m \leq n \leq N} I_{m,n}$. To compute $|E|$, we consolidate the possibly overlapping intervals $I_{m,n}$ into disjoint intervals and sum their lengths.

Since we are interested in maximizing $|E|$, we need only consider those probability measures for which $x_{j+1} - x_j \leq 1$ for all j (separating x_j and x_{j+1} further will only reduce $|E|$). Suppose we fix the value of N and require that $x_1 = 0$ and that each x_j and each a_j be multiples of some small number b. There are then only a finite number of possibities for μ. Using a computer, we can compute $|E|$ for each of these and determine the maximum value of $|E|$ and the measure that gives this value. This is what we mean by a search to a given resolution.

We performed such exhaustive searches for small values of N. Clearly, such a search does not prove anything; a search at a finer resolution may have revealed a maximum that was previously missed. However, a small change in the configuration can only lead to a small change in $|E|$, so for $N \leq 5$, we were able to determine the best configurations with a fair degree of certainty.

For $N = 2$, 3, and 4, the best configurations correspond to those described in the proof of Theorem 2.1, which we refer to as the even configurations of N masses (see Figure 4.1a). The masses are evenly distributed and evenly spaced so that in general $I_{n,n}$ exactly touches $I_{n-1,n}$ and $I_{n,n+1}$. Intervals of the form $I_{m,n}$, where $n - m > 2$, are empty; those for which $n - m = 2$ are negligible because they consist of a single point. In other words, we can compute $|E|$ by considering various single point masses and adjacent pairs of point masses. If we are to increase the value of $|E|$, we must find a configuration where some intervals resulting from the interactions of three or more point masses (intervals of the form $I_{m,n}$, with $n - m \geq 2$) have positive measure. Although such configurations are possible using three or more point masses, they do not maximize $|E|$ when $N = 3$ or 4 because of the large amount of overlap between the various intervals $I_{m,n}$.

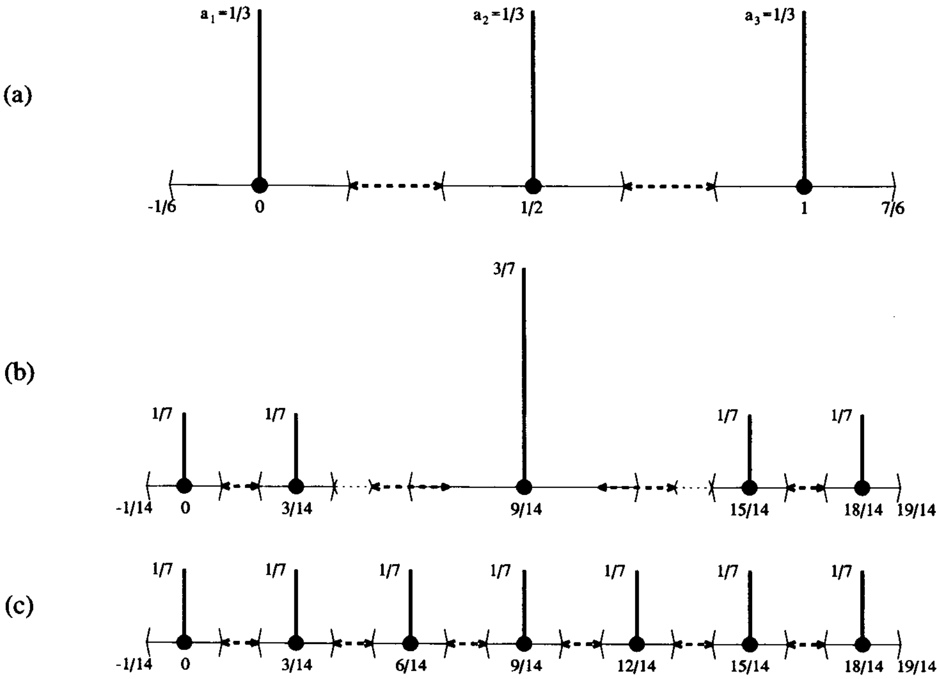


FIGURE 4.1. For three point masses, the optimal configuration (a) is the even configuration. For five masses, the optimal configuration (b) is not even, but covers exactly the same interval as the even configuration of seven masses (c). In these diagrams, the large dots represent the locations of the point masses, the heights of the vertical bars represent their weights, and the horizontal lines represent the various intervals $I_{m,n}$. Solid horizontal lines indicate intervals covered by one point mass ($I_{n,n}$), dashed lines indicate those covered by the interaction of two point masses ($I_{n,n+1}$), and dotted lines indicate those covered by the interaction of three point masses ($I_{n,n+2}$).

When $N = 5$, however, one can improve on the even configuration of Theorem 2.1. The best configuration with five masses is shown in Figure 4.1b. This configuration resembles the even configuration of seven masses, shown in Figure 4.1c, with the three central masses consolidated into a single mass in the center. The two configurations cover the same interval, but in the five-mass configuration, $I_{1,3}$ and $I_{3,5}$ cover part of it, and these are intervals that lie in E because of the interaction of three point masses. Note that the various $I_{m,n}$ are not disjoint in this case. They overlap in several places. Using more point masses, we can construct more complicated configurations resembling this one. With nine point masses, for example, we have a configuration that covers the same interval E as the even configuration of 25 masses. However, we cannot eliminate the overlap between the various $I_{m,n}$ without decreasing the total measure $|E|$. In general, these configurations seem to achieve the same value of $|E|$ as an even configuration involving more point masses. We conjecture that as the number of point masses increases, the best configuration becomes more and more complex, taking advantage of intervals of the form $I_{n,n+k}$ for greater and greater k to cover ever larger intervals. But we believe these intervals are always of the form $[-1/2j, 3/2 - 1/j]$ for some integer j and could therefore be covered by the even configuration of j masses. The limit of $|E|$ as the number of masses increases remains $3/2$. We therefore conjecture that $A(1, 1) = 3/2$.

The limitations of our technique become apparent as N increases. To begin with, even with $N = 4$, it is difficult to prove which configuration is the best because so many different interactions between different groups of point masses must be considered. For $N \geq 5$, the optimal configurations become increasingly complex. Moreover, for $N \geq 5$, several configurations are "local maxima"—that is, while they may not be optimal configurations, they have the property that a small variation of the mass distribution always leads to a decrease in $|E|$. The prospect of proving which configuration is optimal for each positive integer N is not attractive.

For values of N greater than 5, it is computationally expensive even to perform searches at a resolution high enough to determine the optimal configurations with a fair degree of certainty. Each time we add a point mass to our configuration, we add two independent variables (its location and weight) to the system we wish to optimize. The number of possibilities through which we must search grows exponentially. In fact it grows even more quickly because, as N increases, finer searches are necessary to determine the best of several local maxima. These search techniques helped lead to our conjecture that $A(1, 1) = 3/2$, but other techniques will probably be required to prove the conjecture if it is true.

These practical limitations became an even more serious obstacle when we considered point mass distributions in two dimensions. As in one dimension, we can compute $|\{x : M\mu(x) \geq 1\}|$ for any given configuration. To see patterns in two dimensions, however, a larger number of masses are necessary. Five will certainly not suffice. While we found some attractive configurations in two dimensions, we were not able to perform exhaustive searches of large enough sets of configurations to produce any conclusive results. We may be able to make progress on this problem by improving our search techniques.

5. Computational Aspects of the Search ($p > 1$)

For the case of $p > 1$, we can no longer use finite discrete probability measures to approximate arbitrary functions; instead we use step functions. As with the probability measures, we can search over the space of functions such that $\|f\|_p = 1$ by fixing the number and length of the steps (as opposed to point-masses) and again varying their values (heights) by a small "resolution." Again, this reduces our search space to a finite number of functions, given a certain resolution.

Thus, we needed to be able to compute $\|Mf\|_p$ for any given step function f. Suppose f is a step function given as

$$f(x) = \begin{cases} c_j \text{ if } x \text{ is in } [a_j, a_{j+1}) & \text{for } 0 \leq j \leq N-1, \\ 0 & \text{if } x < a_0 \text{ or } x \geq a_N. \end{cases} \tag{5.1}$$

It is possible to compute $\|Mf\|_p$ because, for any x, we need only consider a finite set of values of r as possibilities in the expression of the maximal function—namely, the set of distances $|x - a_j|$ from x to the endpoints of the intervals of the step function. The reason is that as r increases so that $x + r$ and $x - r$ remain in the same interval,

$$\frac{1}{2r}\int_{x-r}^{x+r} f(t)\,dt = a + b/(c + r) \quad \text{for constants } a,\ b \text{ and } c; \tag{5.2}$$

this function is either increasing or decreasing depending on the sign of b, so it assumes its supremum (and thus is equal to $Mf(x)$) at one of the endpoints of the intervals. So, for f a step function as above,

$$Mf(x) = \max\left\{\frac{1}{2r}\int_{x-r}^{x+r} f(t)\,dt : r = |x-a_j|,\ j = 0,\ldots,N\right\}. \tag{5.3}$$

For any of these possible choices of r, it is easy to compute the integral of f over $(x-r, x+r)$. It is the sum of c_j such that (a_j, a_{j+1}) is contained in $[x-r, x+r]$ plus one term of the form $c_k(x+r-a_k)$ or $c_k(a_{k+1}-x+r)$ as the contribution to the integral from an interval (a_k, a_{k+1}) that partially overlaps $[x-r, x+r]$. Thus, when f is a step function, computing $Mf(x)$ for any given x is merely finding the largest of a finite set of easily calculable numbers.

Moreover, in theory it is possible to calculate $\|Mf\|_p$ exactly. From the above, we see that on certain intervals $Mf(x)$ is the maximum of a finite number of rational functions. Thus, we could solve for which of these rational functions are dominant on which subintervals and integrate accordingly. As it turns out, we used a combination of these "exact" techniques and numerical integration to compute $\|Mf\|_p$. For the best combination of computational accuracy, speed, and simplicity, we chose to do numerical integration on the support of the step function and exact integration outside its support.

Suppose f is a step function of N steps as above, with support in $[0, a_N)$. Here is the outline of the strategy for integrating $Mf(x)$ "to the right of the support," over (a_N, ∞) (the strategy is symmetrically similar to the left). For each point in this set, there are the same N "candidates" for $Mf(x)$—those obtained by taking r to be $x - a_j$ for $0 \le j \le N-1$. Note that $\int_{x-r}^{x+r} f(t)\,dt$ is merely $\sum_{i\ge j} c_j$. So, each of the functions in (5.3) is a rational function as in (5.2). By algebraically solving for when pairs of these functions are equal, we can determine which functions are dominant on which subintervals of (a_N, ∞) and then integrate accordingly. This task is made computationally simpler by the fact that we need only look at these functions "in order." If we have determined that the function obtained by taking $r = x - a_k$ is dominant over some interval beginning at a point y, then we need only compare it with the functions obtained by taking successively larger r; i.e., first take $r = x - a_k - 1$, then $r = x - a_k - 2$, and so on, until one of the resulting functions takes over at some point $z > y$.

These computations, though possible on the support of the step function as well, are much more complicated. So instead we used numerical integration by Simpson's method, on the support of the step function, because we could much more easily provide the necessary information for numerical integration, namely, the value of $Mf(x)$ at a number of discrete points.

An example, where this entire procedure is carried out for a step function with two steps, is shown in Figure 5.1.

We initially approached the problem of searching through spaces of step functions in the same manner we approached that of the finite discrete probability measures—we searched through the space of functions f such that $\|f\|_p = 1$, beginning with small numbers of steps and progressing to larger numbers as far as computationally feasible. From the optimal configurations yielded by these searches for small numbers of steps, we hoped to find a pattern pointing toward the ideal configuration in the limit (i.e., as number of steps goes to infinity).

For all our searches involving step functions, we made the support of an "N-step" function $[0, N)$, with the ith step over $[i-1, i)$, and required that the value of each step c_j was a multiple of some fixed small number b. We also noticed early in our investigation that only step functions with an odd number of steps would be useful for our purposes. We reasoned that the ideal configuration in the limit will be a symmetric configuration; with small numbers of steps, only functions with an odd number of steps could exhibit such behavior.

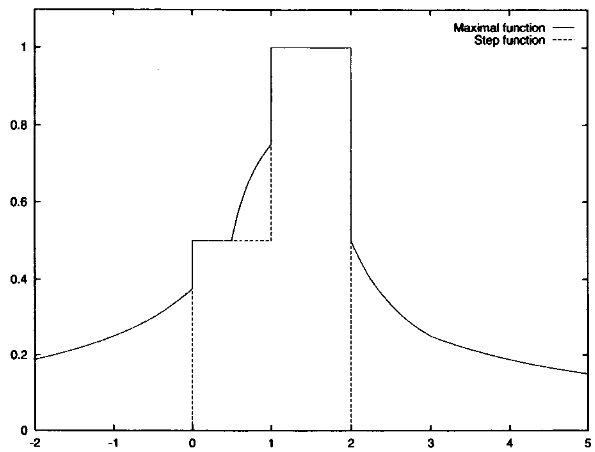


FIGURE 5.1. The maximal function of the step function

$$f(x) = \begin{cases} 0.5 & \text{if } x \text{ in } [0, 1), \\ 1.0 & \text{if } x \text{ in } [1, 2), \\ 0 & \text{otherwise.} \end{cases}$$

For each x, there are three possibilities for r in the expression of $Mf(x)$: $|x|$, $|x-1|$, and $|x-2|$. Each produces a "candidate" for $Mf(x)$. This set of candidates will be identical for points within the following intervals: $(-\infty, 0)$, $(0, 1/2)$, $(1/2, 1)$, $(1, 3/2)$, $(3/2, 2)$, and $(2, +\infty)$. (Note that any step function with steps of measure 1 behaves similarly; there will be a distinct set of candidates for $Mf(x)$ on each half of each step.) For each of these intervals, we listed the candidates for $Mf(x)$ and identified which of these is dominant; i.e., which is actually $Mf(x)$.

Thus, let us turn to the results of our searches through functions with three, five, seven, and nine steps. We first examined the case of $p = 2$; fairly "fine" searches through the spaces of functions with three, five, and seven steps yielded no discernible pattern through the optimal configurations in these cases. However, our experiments with functions of seven steps revealed an interesting phenomenon. There is a much more regular arrangement, whose step heights decrease as one moves outward from the center and which yields a maximal function with p-norm only slightly less than that yielded by the irregular optimal configuration. Our nine-step search revealed that such a symmetric decreasing arrangement is the best configuration with nine steps. This seems also to be the case with eleven steps. These arrangements seem to be the same as those produced by the steepest ascent search as described below. We denote them by $SD(N, p)$.

So, in the case of $p = 2$, it seemed as if $SD(N, p)$, though not the best configuration for small N, became the best configuration once there were "enough" steps. We next tried to see if this behavior held for other p. Working on the idea that the L^p spaces are more "flexible" for smaller p and less so for larger p, we tested functions with $p = 1.5$. As we expected, the $SD(N, p)$ eventually became the optimal configuration, but now such a pattern became the dominant one with only five steps. With three steps, we still saw a "nondecreasing" pattern as the best. But, when we decreased p to 1.25, $SD(N, 1.25)$ was the optimal configuration even with only three steps.

Conversely, our searches with larger p supported the idea that these L^p spaces are less flexible. For example, with $p = 4$, for both nine and eleven steps, $SD(N, 4)$ is not the best configuration. Thus, it seems likely that for any p there is some integer N_p such that if $N \geq N_p$, $SD(N, p)$ is the optimal configuration among step functions with N steps. Moreover, the number of steps needed to see this behavior increases as p increases.

The conjecture that $A(p, 1) = B(p, 1)$ gained support from our work with functions composed of a larger number of steps N. As we increase N, the number of possible functions at any given resolution grows quickly, making exhaustive searches difficult. Such searches become even more difficult because higher resolution is necessary to determine which one of several configurations with similar values of $\|Mf\|_p$ is in fact optimal. We therefore use an alternate approach. One can think of $g = \|Mf\|_p$ as a single-valued function of N independent variables that determine the height of each of the N steps in f. If the function g is sufficiently smooth, one can apply calculus-based search algorithms to find local maxima. We implemented a steepest ascent search, which involves choosing some step function as an initial point, approximating the gradient of g at that point, moving in the direction of the gradient until the function stops increasing, and then recomputing the gradient and repeating the process. The function g is not differentiable everywhere, but it turns out to be smooth enough that the steepest ascent search brings us readily to local maxima.

These local maxima become increasingly numerous as N increases; we have no systematic way to find them all. We noticed, however, that for all but very small values of N, we can achieve the previously noted $SD(N, p)$ by a steepest ascent search beginning with an initial configuration in which the entire weight of the step function is concentrated in the central step. This allows us to produce the configurations we believe are optimal for odd values of N as high as 499.

As N increases, the configurations seem to approach curves reminiscent of functions of the form $f(x) = |x|^{-\alpha}$ for some α. These configurations suggested the approach described in §3. The step functions shown in Figures 5.2a, b and 5.3 represent the best configurations that we were able

(a)

(b)

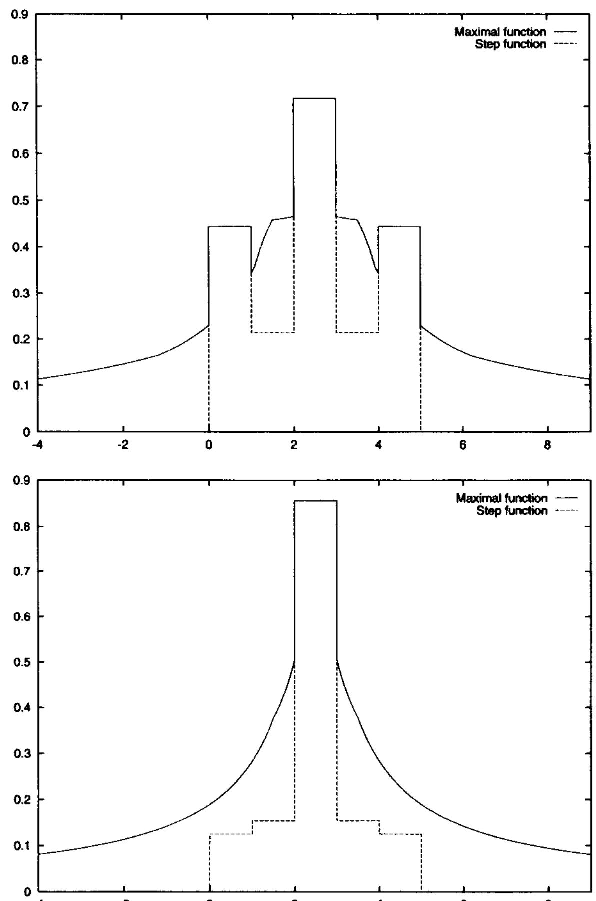


FIGURE 5.2. Optimal configurations with five steps for (a) $p = 2$ and (b) $p = 1.5$. The configuration in (a) is not a symmetric decreasing arrangement, but that in (b) is $SD(5, 1.5)$.

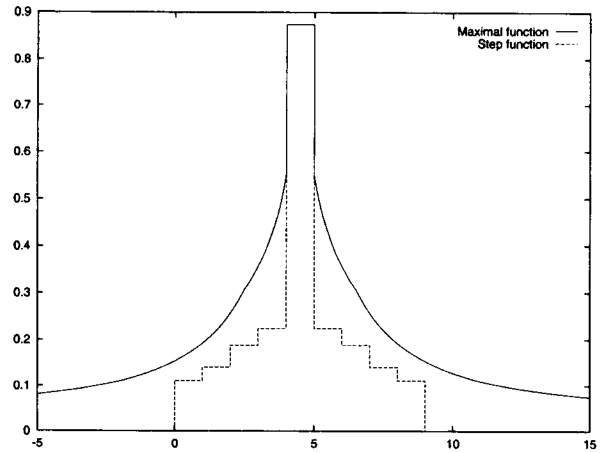


FIGURE 5.3. The optimal configuration with nine steps for $p = 2$ is $SD(9,2)$.

to find with the given numbers of steps. Note, however, that they give values of $\|Mf\|_p$ well below the lower bounds for $A(p, 1)$ given in Theorem 3.2. For $p = 2$, for example, arbitrarily close approximations to the function $f(x) = |x|^{-1/2}$ give values of $\|Mf\|_p$ approaching $1.612\ldots$; while $SD(499, 2)$, shown in Figure 5.4, gives $\|Mf\|_p = 1.459\ldots$. This suggests two possibilities. One is that while the $SD(N, p)$ do not approximate $|x|^{-1/p}$ well in terms of the values they give for $\|Mf\|_p$, they are the best possible approximations with any given number of steps. The other possibility is that, for some given number of steps, we can improve on $SD(N, p)$ by more closely approximating the function $|x|^{-1/p}$. All our attempts to approximate this function with a finite number of steps, however, produced values of $\|Mf\|_p$ lower than that of the $SD(N, p)$ with the same number of steps. We therefore believe that the $SD(N, p)$ are indeed the optimal configurations for large numbers of steps and that a very large number of steps is necessary to obtain a value of $\|Mf\|_p$ close to $B(p, 1)$. While the rate of convergence seems logarithmic, it is faster for lower values of p. Figure 5.5a shows $SD(99, 1.01)$. For comparison, Figure 5.5b shows an approximation to the function $|x|^{-1/1.01}$, which lies in $L^{1.01}$. Note, first, the similarity between the two graphs, and second, the fact that both give values of $\|Mf\|_p$ close to $B(1.01, 1)$. The primary difference between the two graphs is in the

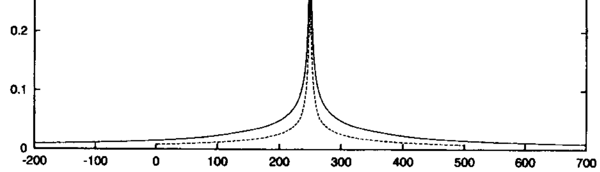
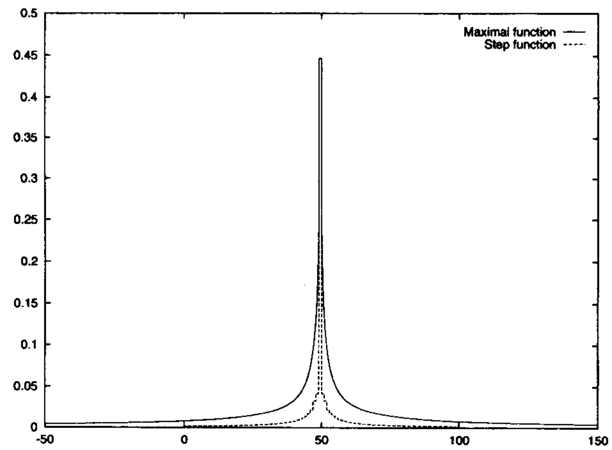


FIGURE 5.4. $SD(499,2)$, which we believe is the optimal configuration with 499 steps for $p = 2$.

region close to the center, where we would expect the greatest error in approximating a function that goes to infinity. For any p, we conjecture that as N increases, the optimal configurations better and better approximate $f(x) = |x|^{-1/p}$. The value for the constant $A(p, 1)$ would then be $B(p, 1)$, as described in Theorem 3.2.

(a)



(b)

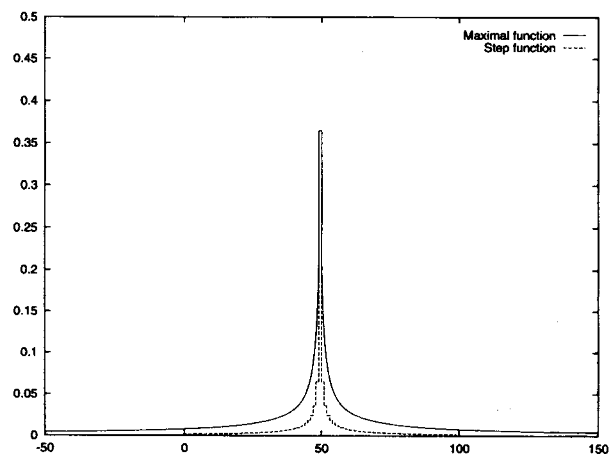


FIGURE 5.5. (a) $SD(99,1.01)$ has $\|Mf\|_{1.01} = 96.295$. (b) This 99-step approximation to $f(x) = x^{-1/1.01}$ was determined by using values of $f(x)$ at regular intervals as step heights and normalizing so that the step function has a 1.01-norm of 1. Here $\|Mf\|_{1.01} = 96.284$, slightly less than the value of $SD(99,1.01)$. Note, however, that the two norms are very close to each other and relatively close to $B(1.01, 1) = 98.277$.

References

[1] Christ, M., and Grafakos, L. (1995). Best constants for two nonconvolution inequalities. *Proc. Amer. Math. Soc.* **123**, 1687–1693.

[2] de Guzman, M. (1981). Real-variable methods in Fourier analysis. *Mathematical Studies* **46**. North-Holland, Amsterdam.

[3] Grafakos, L., and Montgomery-Smith, S. (1995). Best constants for uncentered maximal functions, preprint.

[4] Grafakos, L., Montgomery-Smith, S., and Motrunich, O. (1996). A sharp estimate for the Hardy-Littlewood maximal function, preprint.

[5] Stein, E. M. (1970). *Singular Integrals and Differentiability Properties of Functions.* Princeton University Press, Princeton, NJ.

[6] Stein, E. M. (1982). The development of the square function in the work of A. Zygmund. *Bull. Amer. Math. Soc.* **7**, 359–376.

[7] Stein, E. M. (1993). *Harmonic Analysis: Real-Variable Methods, Orthogonality, and Oscillatory Integrals.* Princeton University Press, Princeton, NJ.

[8] Stein, E. M., and Strömberg, J.-O. (1983). Behavior of maximal functions in $\mathbb{R}^n$ for large n. *Ark. Mat.* **21**, 259–269.

Received July 24, 1995

Dror, Mathematics Department, Rice University, Houston, Texas 77251

Ganguli and Strichartz, Mathematics Department, White Hall, Cornell University, Ithaca, NY 14853