An L_p version of the Beck-Fiala conjecture

Jiří Matoušek*

Department of Applied Mathematics
Charles University
Malostranské nám. 25, 118 00 Praha 1
Czech Republic

Abstract

Beck and Fiala conjectured in 1981 that for any set system $\mathcal{S}$ of maximum degree t on a finite ground set X, a coloring $\chi : X \to \{-1, +1\}$ exists such that $|\chi(S)| = O(\sqrt{t})$ holds for all $S \in \mathcal{S}$, where $\chi(S) = \sum_{x \in S} \chi(X)$. We prove a weaker statement, namely that for any fixed $p \geq 1$, a coloring χ exists such that the pth degree average of $|\chi(S)|$ over $S \in \mathcal{S}$ is $O(\sqrt{t})$. The result also holds if each set is assigned a nonnegative real weight and the pth degree average is taken with these weights (with χ depending on the weights).

1 Introduction

Let X be a finite set and $\mathcal{S} \subseteq 2^X$ be a family of subsets of X. By a *coloring* we mean a mapping $\chi : X \to \{-1, +1\}$. The *discrepancy* of $\mathcal{S}$, denoted by $\mathrm{disc}(\mathcal{S})$, is the minimum, over all colorings χ, of

$$\mathrm{disc}(\mathcal{S}, \chi) = \max_{S \in \mathcal{S}} |\chi(S)| \; , \tag{1}$$

where we use the shorthand $\chi(S)$ for $\sum_{x \in S} \chi(x)$. A very significant result in discrepancy theory is a theorem of Beck and Fiala [4]: *The discrepancy of any*[1] *set system $\mathcal{S}$ is at most $2t - 1$, where t denotes the*

*Research supported by Czech Republic Grant GAČR 0194/1996 and by Charles University grants No. 193,194/1996.

[1] We only consider set systems with a finite ground set in this paper.

maximum degree of $\mathcal{S}$, *i.e.* $t = \max_{x\in X} |\{S \in \mathcal{S}; x \in S\}|$. Beck and Fiala conjectured that the upper bound can be improved to $C\sqrt{t}$, with an absolute constant C (this, if true, is the best possible bound one can hope for, since set systems exist with maximum degree t and discrepancy of the order $\sqrt{t}$). A more detailed discussion (heuristic reasons for the validity of the conjecture, related conjectures, and background information) can be found in [9] (see also [1], [5]).

Here we prove an analogue of the Beck-Fiala conjecture for a weaker notion of discrepancy, where instead of the maximum (worst set in $\mathcal{S}$) as in (1) one takes an average over the sets of $\mathcal{S}$. Namely, for a number $p \in [1,\infty)$ we define

$$\mathrm{disc}_p(\mathcal{S},\chi) = \left(\frac{1}{|\mathcal{S}|}\sum_{S\in\mathcal{S}} |\chi(S)|^p\right)^{1/p};$$

$\mathrm{disc}_p(\mathcal{S})$ is then defined as the minimum of $\mathrm{disc}_p(\mathcal{S},\chi)$ over all colorings χ. More generally, we can consider a weight function $w : \mathcal{S} \to (0,\infty)$, and set

$$\mathrm{disc}_{p,w}(\mathcal{S},\chi) = \left(\frac{1}{w(\mathcal{S})}\sum_{S\in\mathcal{S}} w(S)|\chi(S)|^p\right)^{1/p}.$$

Srinivasan [10] proved that under the hypothesis of the Beck-Fiala theorem, we have $\mathrm{disc}_p(\mathcal{S}) = O(t^{3/4}\log t)$ for each p, with the constant of proportionality depending on p. Here we establish an asymptotically tight bound:

Theorem 1.1 *For each* $p \in [1,\infty)$ *there exists a constant* C_p *such that for any set system* $\mathcal{S}$ *of maximum degree* t *and for any weight function* $w : \mathcal{S} \to (0,\infty)$, *we have*

$$\mathrm{disc}_{p,w}(\mathcal{S}) \le C_p\sqrt{t}\,.$$

The proof follows a known technique, but the adjustment of its numerous parameters seems to require some subtlety for obtaining the tight bound.

Theorem 1.1 can be regarded as a certain support for believing in the Beck-Fiala conjecture itself. On the other hand, situations are known in geometric discrepancy theory where the pth degree average discrepancy is provably smaller than the worst-case discrepancy (for instance, for axis-parallel rectangles in the plane; see e.g., [3]). In combinatorial discrepancy, a natural case where $\mathrm{disc}_p(.)$ is asymptotically smaller than $\mathrm{disc}(.)$ is noted in [6]. Also, the method of our proof of Theorem 1.1 seems to be inadequate for attacking the Beck-Fiala conjecture.

2 Summary of the entropy method

Let us define a *partial coloring* to be a mapping $\chi : X \to \{0, +1, -1\}$, and a let *substantial* partial coloring be a partial coloring χ with $\chi(x) \neq 0$ for at least $|X|/2$ points $x \in X$. The points x with $\chi(x) = 0$ will be called *uncolored* by χ.

A method invented, in its original form, by Beck [2] and further elaborated by Beck, Spencer, Boppana, and the author (see e.g., [8], [7] and references therein) allows one prove the existence, under suitable assumptions, of a substantial partial coloring of a given set system $\mathcal{S}$. In more recent applications, the method is referred to as the *entropy method*, since technically it amounts to estimating the entropy of certain random variables associated with the considered set system. Here we need not introduce the entropy concept and explain the way the required partial coloring is produced. The following proposition, summarizing results of calculations made in [7], can be used as a black box:

Proposition 2.1 (Entropy method) *Let $\mathcal{S}$ be a set system on an n-point set X, and let a number $\Delta_S > 0$ be given for each $S \in \mathcal{S}$. Suppose that*

$$\sum_{S \in \mathcal{S}} h\left(\frac{\Delta_S}{\sqrt{|S|}}\right) \leq \frac{n}{5} \tag{2}$$

holds, where the function $h(\lambda)$ can be estimated by

$$h(\lambda) \leq g(\lambda) = \begin{cases} Ke^{-\lambda^2/9} & \text{if } \lambda > 0.1 \\ K\ln(\lambda^{-1}) & \text{if } \lambda \leq 0.1 \end{cases}$$

with an absolute constant K. Then there exists a substantial partial coloring $\chi : X \to \{\pm 1\}$ such that $|\chi(S)| \leq \Delta_S$ for all $S \in \mathcal{S}$.

As a warm-up example of an application of this result, let us re-prove an upper bound for the discrepancy under the conditions of the Beck-Fiala theorem. This bound is the best known one for the case when $|X|$ is less than exponential in t. It was proved by Srinivasan [10], improving previous slightly weaker results of Beck and Spencer.

Theorem 2.2 ([10]) *Let $\mathcal{S}$ be a set system of maximum degree t on an n-point set X. Then* $\operatorname{disc}(\mathcal{S}) = O(\log n\sqrt{t})$*, with an absolute constant of proportionality.*

Proof. Using Proposition 2.1, we prove that any set system $\mathcal{S}$ of maximum degree t has a substantial partial coloring with discrepancy $O(\sqrt{t})$. Having established this, the theorem is proved by a standard iteration argument. Namely, given a set system $\mathcal{S}$ on an n-point set X, we consider a substantial partial coloring χ_0 of X with discrepancy $O(\sqrt{t})$. We let X_1 be the set of at most $n/2$ points uncolored by χ_0, and let $\mathcal{S}_1$ be $\mathcal{S}$ restricted to X_1. The maximum degree of $\mathcal{S}_1$ is again at most t, and hence a substantial partial coloring χ_1 of X_1 exists with $\operatorname{disc}(\mathcal{S}_1, \chi_1) = O(\sqrt{t})$. The partial colorings χ_0 and χ_1 together leave at most $n/4$ points uncolored, and for these we produce another partial coloring χ_2, etc. After $O(\log n)$ iterations of this partial coloring step, we end up with at most a constant number of points still uncolored, and these can be colored arbitrarily. The resulting coloring of X has discrepancy $O(\log n\sqrt{t})$.

To get a substantial partial coloring χ for a set system $\mathcal{S}$ with maximum degree t, we apply Proposition 2.1 with $\Delta_S = C\sqrt{t}$ for a large enough constant C; all we need to do is checking the condition (2). To this end, we observe that for any given size s, $1 \leq s \leq n$, $\mathcal{S}$ contains at most nt/s sets of size $\geq s$, where n is the size of the ground set. In order to sum up the contributions of sets of various sizes conveniently, we put $s_i = 2^{-i}100C^2t$ (i an integer, possibly also negative), and we let $\mathcal{S}_i$ consist of the sets of $\mathcal{S}$ whose size lies in the interval $(\frac{s_i}{2}, s_i]$. Then

$$\sum_{S\in\mathcal{S}} h\left(\frac{C\sqrt{t}}{\sqrt{|S|}}\right) \leq \sum_i |\mathcal{S}_i| h\left(\frac{C\sqrt{t}}{\sqrt{s_i}}\right) \leq \sum_i \frac{2nt}{s_i} h\left(\tfrac{2^{i/2}}{10}\right) \leq$$

$$\frac{K}{50C^2}\, n \left[\sum_{2^i\geq 1} 2^i e^{-2^i/900} + \sum_{2^i<1} 2^i \ln(10.2^{-i/2})\right].$$

Clearly, both sums in brackets are bounded by absolute constants, and hence this bound can be pushed below $\frac{n}{5}$ by taking C large enough. □

[6] J. Matoušek. On discrepancy bounds via dual shatter function. KAM Series (Tech. Report) 94–273, Charles University, Prague, 1994. Also to appear in *Mathematika*.

[7] J. Matoušek and J. Spencer. Discrepancy in arithmetic progressions. *J. Amer. Math. Soc.*, 9:195–204, 1996.

[8] J. Spencer. Six standard deviations suffice. *Trans. Amer. Math. Soc.*, 289:679–706, 1985.

[9] J. Spencer. *Ten lectures on the probabilistic method.* CBMS-NSF. SIAM, 1987.

[10] A. Srinivasan. Improving the discrepancy bound for sparse matrices: better approximations for sparse lattice approximation problems. *Proc. ACM Symposium on Discrete Algorithms*, to appear, 1997.

3 The case of equal weights

In this section we prove a particular case of Theorem 1.1, namely that $\mathrm{disc}_p(\mathcal{S}) \le C_p\sqrt{t}$ holds for any set system $\mathcal{S}$ of maximum degree t on a ground set X with $m = |\mathcal{S}| \ge n = |X|$. Hopefully, this might help in reading the (much messier) proof of the general case (arbitrary set weights).

The desired coloring χ will be produced similarly as in the above proof of Theorem 2.2. A partial coloring step is iterated until only negligibly few uncolored points remain. The choice of the numbers Δ_S is slightly different, however, as will be specified below.

Notational convention. Instead of numbering the iterations of the partial coloring step consecutively, we assign the index k to an iteration of the partial coloring step such that the number of yet uncolored points at the beginning of the iteration is in the interval $(\frac{n}{2^{k+1}}, \frac{n}{2^k}]$. Since in each step, at least half of the currently uncolored points becomes colored, no two steps receive the same indices, but the indices need not form a contiguous sequence. The symbol $\sum_k$ in the sequel means the summation over the indices of all iterations performed.

Let X_k be the set of yet uncolored points at the beginning of step with index k, let $n_k = |X_k| \in \left(\frac{n}{2^{k+1}}, \frac{n}{2^k}\right]$ be their number, and let χ_k denote the partial coloring produced in that step. As in the proof of Theorem 2.2, put $s_i = 2^{-i}100C^2t$ for a (possibly negative) integer i, and let $\mathcal{S}_{k,i}$ be the system of all sets in $\mathcal{S}$ restricted to X_k whose size lies in the interval $(\frac{s_i}{2}, s_i]$.

For the sets $S \in \mathcal{S}_{k,i}$, we put

$$\Delta_S = \Delta_i = C\sqrt{t}\,\varphi(i)\,,$$

where $\varphi(i) = 2^{-|i|/4p}$ is an auxiliary "bump-like" function (it is 1 for $i = 0$ and decreases geometrically but slowly enough for i going away from 0 in both directions). Hence we require that $|\chi_k(S)| \le \Delta_i$ for all $S \in \mathcal{S}_{k,i}$.

First need to show that such partial colorings can indeed be enforced, i.e. to check the condition (2) in Proposition 2.1. This is fairly similar to the calculation in the proof of Theorem 2.2, so we omit this part and pass directly to verification that $\mathrm{disc}_p(\mathcal{S},\chi) = O(\sqrt{t})$ holds for the resulting coloring.

For a vector v indexed by the sets of $\mathcal{S}$, write $\|v\|_p = \left(\frac{1}{m}\sum_{S\in\mathcal{S}} |v(S)|^p\right)^{1/p}$

(where $m = |\mathcal{S}|$). With this notation, we prove the estimate

$$\|\chi_k\|_p \le \alpha(k)\sqrt{t}$$

with an auxiliary function $\alpha(k)$ such that $\sum_{k=0}^{\infty} \alpha(k) = O(1)$. Then we get

$$\mathrm{disc}_p(\mathcal{S}, \chi) = \|\chi\|_p \le \sum_k \|\chi_k\|_p = O(\sqrt{t})$$

as claimed.

We have

$$\|\chi_k\|_p^p = \frac{1}{m} \sum_i \Delta_i^p |\mathcal{S}_{k,i}| \le \frac{(C\sqrt{t})^p}{m} \sum_i \varphi(i)^p \min\left(m, \frac{2n_k t}{s_i}\right) \le$$

$$(C\sqrt{t})^p \sum_i \varphi(i)^p \min\left(1, 2^i \frac{n}{C_1 2^k m}\right)$$

where $C_1 = 50C^2$. Let $i_0 = i_0(k) = \log_2 \frac{C_1 2^k m}{n}$. Since we assume $m \ge n$, we get that $i_0 \ge k - O(1)$. We can re-write

$$\sum_i \varphi(i)^p \min\left(1, 2^i \frac{n}{C_1 2^k m}\right) = \sum_{i \ge i_0} \varphi(i)^p + \sum_{i < i_0} \varphi(i)^p 2^i \frac{n}{C_1 2^k m} =$$

$$\sum_{i \ge i_0} \varphi(i)^p + O(\varphi(i_0)^p) = \sum_{i \ge k} \varphi(i)^p + O(\varphi(k)^p) = O(\varphi(k)^p)\,.$$

Therefore $\|\chi_k\|_p = O(\sqrt{t}\,\varphi(k))$ and the proof of the considered particular case of Theorem 1.1 is complete. □

4 The general case

Here we prove Theorem 1.1. For a more convenient notation, we will use $f \ll g$ as a shorthand for $f = O(g)$. The implicit constant in the $O(.)$ notation may depend on p and other parameters declared as constant.

As in the preceding section, $p \ge 1$ is a fixed number, $\mathcal{S} \subseteq 2^X$ is the considered set system of maximum degree at most t, $n = |X|$, and $m = |\mathcal{S}|$. Moreover, we consider a real weight function $w : \mathcal{S} \to (0, \infty)$. By a suitable re-scaling, we may assume that $w(\mathcal{S}) = \sum_{S \in \mathcal{S}} w(S) = m$.

Now the first sum doesn't depend on the set weights, and a precisely the same calculation as in the unweighted case (with n_k instead of n) shows that it can be made smaller than $\frac{n_k}{10}$ by setting C large enough. So it remains to deal with the second sum:

$$\sum_{q;\, \psi(k,q) < 0.1} m_q . 2K \ln \frac{1}{\psi(k,q)} \ll \sum_{q;\, m_q 2^{\delta(q)} < n_k/10C} m_q \ln \frac{n_k}{C m_q 2^{\delta(q)}} \le$$

$$\sum_j 2^{j+1} \sum_{q \in Q(j)} \max\left(\log_2 \frac{n_k}{C 2^{j+\delta(q)}}, 0\right) \le$$

$$\sum_j 2^{j+1} \sum_{\substack{q \in Q(j) \\ j+\delta(q) \le \log_2(n_k/C)}} \left[\log_2 \tfrac{n_k}{C} - j - \delta(q)\right]\,.$$

The inner sum has at most $\log_2 \frac{n_k}{C} - j$ terms, each of value no bigger than $\log_2 \frac{n_k}{C} - j$, and so we get that the sum over j is at most

$$\sum_{j \le \log_2(n_k/C)} 2^j (\log_2 \tfrac{n_k}{C} - j)^2 \ll \frac{n_k}{C}\,,$$

with the constant of proportionality independent of C. Hence also the second sum in (4) can be made smaller than $\frac{n_k}{10}$ by setting C large enough, and Theorem 1.1 is proved. □

References

[1] N. Alon and J. Spencer. *The probabilistic method*. J. Wiley & Sons, 1993.

[2] J. Beck. Roth's estimate on the discrepancy of integer sequences is nearly sharp. *Combinatorica*, 1(4):319–325, 1981.

[3] J. Beck and W. Chen. *Irregularities of distribution*. Cambridge University Press, 1987.

[4] J. Beck and T. Fiala. "Integer making" theorems. *Discr. Appl. Math*, 3:1–8, 1981.

[5] J. Beck and V. Sós. Discrepancy theory. In *Handbook of Combinatorics*, pages 1405–1446. North-Holland, Amsterdam, 1995.

$$\sum_j 2^j 2^{\max Q(j)} \le \sum_q m_q 2^q \ll m\,.$$

Hence $\mathrm{disc}_{p,\bar{w}}(\mathcal{S},\chi) \ll \sqrt{t}$ is verified.

Estimating the entropy. First we check some simple properties of the function $g(\lambda)$ defined in Proposition 2.1.

Lemma 4.1 *The function $g(\lambda)$ is nonincreasing for $\lambda > 0$, and for any $\mu \in (0, 0.1]$ and any $\lambda > 0$ we have $g(\mu\lambda) \le 2g(\mu) + g(\lambda)$.*

Proof of the Lemma. The monotonicity is obvious from definition. The second property is inspired by properties of entropy (whose estimate $g(\lambda)$ really is), but it can be verified purely formally, by discussing few cases. Namely, for $\lambda \le 0.1$ we have $g(\mu\lambda) = g(\mu) + g(\lambda)$. For $\mu\lambda \ge 0.1$, we get $g(\mu\lambda) \le K < 2K\ln 10 < 2g(\mu)$, and finally for the intermediate case $0.1 \le \lambda \le 0.1/\mu$ we get $g(\mu\lambda) = K\ln(\mu^{-1}) + K\ln(\lambda^{-1}) \le 2K\ln(\mu^{-1}) = 2g(\mu)$. □

In the rest of the proof, the constants implicit in the $O(.)$ and $\ll$ notation will be independent of C. To conclude the proof of Theorem 1.1, we need to verify that the condition (2) is satisfied at each step k, which in our case means

$$\sum_q \sum_i |\mathcal{S}_{k,i}^{(q)}|\; g\left(\frac{\Delta_{k,i}^{(q)}}{\sqrt{s_i}}\right) \le \frac{n_k}{5}$$

for all k. By plugging in the expressions for s_i and for $\Delta_{k,i}^{(q)}$, distinguishing the cases $\psi(k,q) \le 0.1$ and $\psi(k,q) > 0.1$, and applying Lemma 4.1, the l.h.s. can be bounded as follows:

$$\sum_q \sum_i |\mathcal{S}_{k,i}^{(q)}|\; g\left(\tfrac{2^{i/2}\varphi(i)\psi(k,q)}{10}\right) \le$$

$$\sum_{q;\,\psi(k,q)\ge 0.1} \; \sum_i |\mathcal{S}_{k,i}^{(q)}|\; g\left(\tfrac{2^{i/2}\varphi(i)}{100}\right) +$$

$$\sum_{q;\,\psi(k,q)<0.1} \; \sum_i |\mathcal{S}_{k,i}^{(q)}| \left[g\left(\tfrac{2^{i/2}\varphi(i)}{10}\right) + 2g(\psi(k,q))\right] \le$$

$$\sum_i |\mathcal{S}_{k,i}|\; g\left(\tfrac{2^{i/2}\varphi(i)}{100}\right) + \sum_{q;\,\psi(k,q)<0.1} |\mathcal{S}^{(q)}|.2g(\psi(k,q))\,. \qquad (4)$$

As a first step, we define a new weight function $\bar{w}$ as follows:

$$\bar{w}(S) = \min\{2^q;\, 2^q \ge w(S),\, q = 0,1,2,\ldots\}\,.$$

Since $\bar{w}(S) \le 2w(S) + 1$ for all $S \in \mathcal{S}$, we have $\bar{w}(\mathcal{S}) \le 3m$. Hence it suffices to prove the existence of a coloring $\chi : X \to \{-1,+1\}$ such that

$$\left(\frac{1}{m}\sum_{S\in\mathcal{S}} \bar{w}(S)|\chi(S)|^p\right)^{1/p} \ll \sqrt{t}\,.$$

The advantage of $\bar{w}$ is that it only attains values of the form 2^q for nonnegative integers q. Let $\mathcal{S}^{(q)}$ consist of the sets $S \in \mathcal{S}$ with weight $\bar{w}(S) = 2^q$.

The coloring χ will again be obtained by an iteration of a partial coloring step as in the previous proofs. From the proof in the preceding section, we preserve the indexing of the iterations, the notation X_k (uncolored points before iteration k), $n_k = |X_k| \in (\frac{n}{2^{k+1}}, \frac{n}{2^k}]$, and the set size thresholds $s_i = 2^{-i}100C^2t$. We let $\mathcal{S}_k^{(q)}$ be the restriction of the system $\mathcal{S}^{(q)}$ on X_k (with empty sets deleted), and $\mathcal{S}_{k,i}^{(q)}$ are the sets of $\mathcal{S}_k^{(q)}$ whose size is in the interval $(\frac{s_i}{2}, s_i]$.

The difference compared to the previous proof is that this time the maximum allowed discrepancy of a set $S \in \mathcal{S}$ under χ_k will also depend on k and on the weight of S (besides the size of S as in the proof in the preceding section). Namely, for $S \in \mathcal{S}_{k,i}^{(q)}$, we set

$$\Delta_S = \Delta_{k,i}^{(q)} = C\sqrt{t}\,\varphi(i)\psi(k,q)\,,$$

where C is a large enough constant, $\varphi(i) = 2^{-|i|/4p}$ is as in the preceding proof, and $\psi(k,q)$ is another auxiliary function.

Before defining $\psi(q,k)$ formally, let us try to indicate its intended meaning. Because of the factor $\varphi(i)$, the only sets whose discrepancy under χ_k may be close to $\sqrt{t}$ are those with $i \approx 0$, i.e. of size about t. For simplicity, let us consider only sets of this "critical" size t for a moment. If we have $m \ge n$ sets, all of weight 1, as in the preceding section, the discrepancy $\mathrm{disc}_p(\mathcal{S},\chi_k)$ is an average of $m \ge n$ numbers. At step k, the number of sets of size t is at most $n_k \approx n/2^k$, and hence the contribution of size-t sets to the pth degree average discrepancy $\mathrm{disc}_p(\mathcal{S},\chi_k)$ starts at $O(\sqrt{t})$ at step $k = 0$ and then decreases geometrically with k. On the other hand, if the set weights are arbitrary, $\mathcal{S}$ may contain some small number (much smaller than n) of sets of size t whose weight is a large fraction of the total weight. The entropy method gives us no control which points get colored by a partial coloring. Hence it might possibly happen that for each of these few critical size-t sets, only about $\sqrt{t}$ points are colored in each step, while their discrepancy also increases by about $\sqrt{t}$ in each such step. But we observe that the number of sets making this kind of troubles at step k must be considerably smaller than n_k, and hence we can put a more strict limit on their discrepancy under χ_k while preserving the condition(2). To make this idea work in all cases, we need a bit complicated definition of $\psi(k,q)$.

Let $m_q = |\mathcal{S}^{(q)}|$ be the number of sets of weight 2^q, and for $j = 0,1,2,\ldots,\lfloor \log_2 m \rfloor$, let us put

$$Q(j) = \{q;\, 2^j \le m_q < 2^{j+1}\}\,.$$

Define j_q as the j with $q \in Q(j)$, and further set

$$\delta(q) = |\{q' \in Q(j_q);\, q' \ge q\}|\,.$$

Now we are ready to define $\psi(k,q)$:

$$\psi(k,q) = \min\left(1, \frac{Cm_q 2^{\delta(q)}}{n_k}\right)\,.$$

This finishes the definition of the discrepancy limits $\Delta_{k,i}^{(q)}$; it remains to check that the condition (2) holds and thus the required partial colorings χ_k satisfying the above-defined limits exist, and that the resulting coloring χ has $\mathrm{disc}_{p,\bar{w}}(\mathcal{S},\chi) = O(\sqrt{t})$. We begin with the latter calculation.

Discrepancy calculation. We need to show

$$\sum_{S\in\mathcal{S}} \bar{w}(S)|\chi(S)|^p \ll m(\sqrt{t}\,)^p\,.$$

The left-hand side can be re-written to

$$\sum_{q\ge 0} 2^q \sum_{S\in\mathcal{S}^{(q)}} |\chi(S)|^p \le \sum_{q\ge 0} 2^q \left[\sum_k \left(\sum_{S\in\mathcal{S}_k^{(q)}} |\chi_k(S)|^p\right)^{1/p}\right]^p. \qquad (3)$$

For the innermost sum, which we denote by $A(k,q)$, we get

$$A(k,q) = \sum_{S\in\mathcal{S}_k^{(q)}} |\chi_k(S)|^p \le \sum_i |\mathcal{S}_{k,i}^{(q)}| \left(\Delta_{k,i}^{(q)}\right)^p \ll$$

$$(\sqrt{t}\,)^p \psi(k,q)^p \sum_i \min\left(m_q, \frac{2n_k t}{s_i}\right) \varphi(i)^p \le$$

$$(\sqrt{t}\,)^p \psi(k,q)^p m_q \sum_i \min\left(1, 2^i \frac{n_k}{C_1 m_q}\right) \varphi(i)^p$$

with $C_1 = 50C^2$. Similarly to the proof in the unweighted case, we split the last sum into two parts at $i = i_0 = i_0(k,q) = \log_2(\frac{C_1 m_q}{n_k})$. The terms with $i < i_0$ decrease geometrically with decreasing i, and so we get

$$A(k,q) \ll (\sqrt{t}\,)^p \psi(k,q)^p m_q \sum_{i\ge i_0} \varphi(i)^p\,.$$

Now if $i_0 \le 0$, i.e. the summation range contains 0, the sum is $O(1)$, and otherwise it is $\ll \varphi(i_0)^p \ll \left(\frac{n_k}{C_1 m_q}\right)^{1/4}$.

Considering the second innermost sum in (3), we thus have

$$\frac{1}{\sqrt{t}} \sum_k A(k,q)^{1/p} \ll m_q^{1/p} \sum_k \psi(k,q) \min\left[1, \left(\frac{n_k}{C_1 m_q}\right)^{1/4p}\right] =$$

$$m_q^{1/p} \sum_k \min\left(1, \frac{Cm_q 2^{\delta(q)}}{n_k}\right) \min\left[1, \left(\frac{n_k}{C_1 m_q}\right)^{1/4p}\right]\,.$$

The value of the last sum over k is proportional to the number of its addends that are close to 1 (for k getting smaller, the first min decreases geometrically, and for k getting larger, the second min does). The number of such addends is $O(\delta(q))$, and hence

$$\sum_q 2^q \left[\sum_k A(k,q)^{1/p}\right]^p \ll (\sqrt{t}\,)^p \sum_q 2^q m_q \delta(q)^p\,.$$

Further

$$\sum_q 2^q m_q \delta(q)^p \le \sum_j 2^{j+1} \sum_{q\in Q(j)} 2^q \delta(q)^p \le \sum_j 2^{j+1} \sum_{\ell=1}^{|Q(j)|} 2^{\max Q(j)-\ell+1} \ell^p \ll$$