

Approximating Fair Clustering with Cascaded Norm Objectives

Eden Chlamtác*

Yury Makarychev†

Ali Vakilian‡

Abstract

We introduce the (p, q) -Fair Clustering problem. In this problem, we are given a set of points P and a collection of different weight functions W . We would like to find a clustering which minimizes the ℓ_q -norm of the vector over W of the ℓ_p -norms of the weighted distances of points in P from the centers. This generalizes various clustering problems, including Socially Fair k -Median and k -Means, and is closely connected to other problems such as Densest k -Subgraph and Min k -Union.

We utilize convex programming techniques to approximate the (p, q) -Fair Clustering problem for different values of p and q . When $p \geq q$, we get an $O(k^{(p-q)/(2pq)})$, which nearly matches a $k^{\Omega((p-q)/(pq))}$ lower bound based on conjectured hardness of Min k -Union and other problems. When $q \geq p$, we get an approximation which is independent of the size of the input for bounded p, q , and also matches the recent $O((\log n / (\log \log n))^{1/p})$ -approximation for (p, ∞) -Fair Clustering by Makarychev and Vakilian (COLT 2021).

1 Introduction

Clustering is one of the fundamental problems in various areas including theoretical computer science, machine learning and operations research. Due to its broad range of applications, different researchers have considered different *cluster models*. A typical model in theoretical computer science is the *centroid model* which contains the classic k -Median, k -Means, k -Center and more generally, k -Clustering with ℓ_p -norm objective. In k -Clustering with ℓ_p -norm objective, we are given a set of points in a metric space and the goal is to find a set of k centers C so as to minimize the total ℓ_p distance of points to C . These problems are all known to be NP-hard and admit various efficient approximation algorithms [Gonzalez, 1985, Hochbaum and Shmoys, 1985, Charikar et al., 2002, Kanungo et al., 2004, Gupta and Tangwongsan, 2008, Li and Svensson, 2016, Ahmadian et al., 2020].

Recently, clustering has been widely studied under fairness constraints to prevent or alleviate the bias and discrimination in the solution constructed by the existing algorithms [Chierichetti et al., 2017, Ahmadian et al., 2019, Bera et al., 2019, Bercea et al., 2019, Backurs et al., 2019, Schmidt et al., 2019, Huang et al., 2019, Kleindessner et al., 2019, Chen et al., 2019, Micha and Shah, 2020, Jung et al., 2020, Mahabadi and Vakilian, 2020, Kleindessner et al., 2020, Brubach et al., 2020]. In particular, the notion we consider in this paper is inspired by the notion of *Socially Fair Clustering* introduced by Abbasi et al. [2021], Ghadiri et al. [2021]. In Socially Fair Clustering, points in the input belong to different groups and the goal is to pick k centers so as to minimize the maximum clustering cost incurred to any of the different (demographic) groups in the input. We note that the objective of Socially Fair Clustering was also previously examined by Anthony et al. [2010] in the context of *Robust k -Clustering* where a set of possible scenarios (i.e., a set of clients) are given in the input and the goal is to find a set of centers which is a good solution for all scenarios.

Here, we consider a generalization of both Socially Fair Clustering and Robust Clustering which is denoted as (p, q) -Fair Clustering and is formally defined as below. Besides the fact that (p, q) -Fair Clustering generalizes various known problems, the new formulation is a generalization of the classic k -Clustering problem and understanding the complexity of the problem under different values of p and q is of interest in itself.

DEFINITION 1.1. ((p, q) -Fair Clustering) An instance \mathcal{I} consists of a metric space $([m], d)$ on m points, n different groups of points which are represented by non-negative weight functions w_1, \dots, w_n (for each $i \in [n]$, $w_i : [m] \rightarrow \mathbb{R}_{\geq 0}$) and a target number of centers k . For each $i \in [n]$, the ℓ_p -cost of group i w.r.t. a set of centers

*Department of Computer Science, Ben-Gurion University. The work was done while the author was visiting and supported in part by TTIC. Email: chlamtac@cs.bgu.ac.il

†Toyota Technological Institute at Chicago (TTIC). Supported by NSF awards CCF-1718820, CCF-1955173, and CCF-1934843. Email: yury@ttic.edu

‡Toyota Technological Institute at Chicago (TTIC). Supported by NSF award CCF-1934843. Email: vakilian@ttic.edu

$C \subseteq [m]$ is defined as

$$(1.1) \quad \text{cost}_p^{\mathcal{I}}(C, w_i) = \left(\sum_{j=1}^m w_i(j) \cdot d(j, C)^p \right)^{1/p},$$

where $d(j, C) = \min_{j' \in C} d(j, j')$. Moreover, the (ℓ_p, ℓ_q) -cost of a set of k centers $C \subseteq [m]$ is defined as the ℓ_q -norm of the vector consisting the ℓ_p -cost of the groups w.r.t. C is defined as

$$(1.2) \quad \text{cost}_{p,q}^{\mathcal{I}}(C) = \left(\sum_{i=1}^n \text{cost}_p(C, w_i)^q \right)^{1/q} = \left(\sum_{i=1}^n \left(\sum_{j=1}^m w_i(j) \cdot d(j, C)^p \right)^{q/p} \right)^{1/q}$$

In (p, q) -Fair Clustering, the goal is to find a set of k centers $C \subseteq [m]$ so as to minimize $\text{cost}_{p,q}^{\mathcal{I}}(C)$.

REMARK 1.1. In the above definition, we may omit the superscript \mathcal{I} indicating the instance relative to which we compute the cost when it is clear from the context.

By setting $q = \infty$, (p, q) -Fair Clustering captures Socially Fair Clustering with ℓ_p -cost. Moreover, compared to Socially Fair Clustering, the formulation of (p, q) -Fair Clustering allows for a “relaxed” way of enforcing the fairness requirement. In particular, by varying the value of q from p to ∞ , the objective interpolates between the objective of classic k -Clustering with ℓ_p -cost and that of Socially Fair Clustering with ℓ_p -cost.¹ Depending on how accurate the group membership information is, the user can set the value of q accordingly and aim for a clustering with a “reasonable” fairness constraint w.r.t. the extracted group membership information: the more accurate the group membership information gets, the higher value can be assigned to q .

We remark that the special case of $(\infty, 1)$ -Fair Clustering was previously studied by [Anthony et al. \[2010\]](#) under the name of *Stochastic k -Center* where the intuition is that we have a potential set of (client) scenarios and know that one of them is likely to happen but do not know which one. [Anthony et al.](#) proved that Stochastic k -Center is as hard to approximate as the Densest k -Subgraph. In particular, $(\infty, 1)$ -Fair Clustering with 0-1 weight functions can be seen to be equivalent to the Min s -Union problem (a generalization of Densest k -Subgraph) [[Chlamtáč et al., 2018, 2017](#)], in which we are given a collection of m sets and an integer $s \in [m]$ and the goal is to choose s sets from the input whose union has minimum cardinality. In addition to the hardness result for Stochastic k -Center, [Anthony et al.](#) also provided an $O(\log m)$ -approximation for the Stochastic k -Median problem which is equivalent to $(1, 1)$ -Fair Clustering with arbitrary weight functions w_1, \dots, w_n .

In a different direction, [Goyal and Jaiswal \[2021\]](#) designed constant factor approximation FPT algorithms for Socially Fair k -Median and k -Means and provided hardness results as well.

REMARK 1.1. [Chakrabarty and Swamy \[2019\]](#) observed that a slightly modified variant of standard existing algorithms for k -Median (e.g., [Charikar et al. \[2002\]](#)) can be applied to obtain a constant factor approximation algorithm for the more general (weighted) k -Clustering with ℓ_p -cost (for $p \in [1, \infty)$). This can also be achieved by applying the algorithm of [[Makarychev and Vakilian, 2021](#)] for (p, ∞) -Fair Clustering when the number of groups is one.

1.1 Our Results and Techniques. In this paper, we design approximation algorithms for (p, q) -Fair Clustering for all values of $p, q \in [1, \infty)$. The following theorems are our main contributions in this work.

THEOREM 1.1. ($p \geq q$) *In the regime $p \geq q$, there exists a polynomial time algorithm that computes an $O\left(k^{\frac{p-q}{2pq}}\right)$ -approximation for (p, q) -Fair Clustering.*

¹We remark that in Socially Fair Clustering with ℓ_p -cost as defined by [Makarychev and Vakilian \[2021\]](#), the goal is to minimize the maximum over all groups i of $\text{cost}_p(C, w_i)^p$. However, in (p, ∞) -Fair Clustering the goal is to minimize the maximum over all groups i of $\text{cost}_p(C, w_i)$. Note that an α -approximation algorithm for (p, ∞) -Fair Clustering implies an α^p -approximation guarantee for Socially Fair Clustering with ℓ_p -cost and vice versa.

This approximation may be nearly optimal in the following sense. As we show in Appendix B, assuming certain hardness conjectures for Min s -Union and related problems, it is in fact not possible to get an algorithm for (p, q) -Fair Clustering in the regime $p > q$ with an approximation factor less than $m^{\Omega((p-q)/(pq))}$ (note that $k \leq m$). See Theorem B.1 for details. Moreover, when $k \ll m$ our algorithm achieves a much better approximation which is independent of m .

THEOREM 1.2. *($p \leq q$) In the regime $p \leq q$, there exists a polynomial time algorithm that computes an $O\left(\left(\frac{q}{\ln(1+q/p)}\right)^{1/p}\right)$ -approximation for (p, q) -Fair Clustering.*

Note that (p, p) -Fair Clustering (i.e., $p = q$) with 0-1 weight functions is equivalent to the classic k -Clustering with ℓ_p -cost. In this setting, Theorem 1.2 gives a constant factor approximation which is essentially optimal. This bound implies an $O\left(\left(\frac{\log n}{\log \log n}\right)^{1/p}\right)$ -approximation for (p, ∞) -Fair Clustering², which matches the recent approximation algorithm of [Makarychev and Vakilian, 2021] and the hardness result of [Bhattacharya et al., 2014]. Thus, for any value of p , the approximation guarantee of Theorem 1.2 smoothly interpolates between the optimal approximation bounds for the previously studied special cases of $q = \infty$ and $q = p$.

Finally, we remark that for any p, q it is possible to get an $O(n^{1/p-1/q})$ approximation by approximating the outer ℓ_q -norm in the formulation of (p, q) -Fair Clustering with ℓ_p -norm and solving the obtained instance of (p, p) -Fair Clustering with one of the existing constant factor approximation algorithms of k -Clustering with ℓ_p -cost.

Overview of our algorithms. At a high-level, our approach is to solve a convex programming relaxation of (p, q) -Fair Clustering and then round the fractional solution to get an approximate integral solution. However, even coming up with an efficient convex program of the problems is non-trivial and introducing such a relaxation is one of our contributions in this paper. While a generalization of the standard LP relaxations for clustering problems (e.g., k -Median [Charikar et al., 2002]) results in a natural convex programming relaxation of (p, q) -Fair Clustering when $p \leq q$, it is less clear how to even come up with a valid convex program relaxation of (p, q) -Fair Clustering for the other scenario, $p \geq q$. In particular, the natural constraint we need to add is not convex when $p > q$. In the $p \leq q$ regime, we cannot use the natural convex relaxation, or even the stronger relaxation which generalizes the LP-relaxation for (p, ∞) -Fair Clustering used by Makarychev and Vakilian, since even the latter has a simple $\Omega(n^{(q-p)/q^2})$ integrality gap construction. To overcome this polynomial lower-bound, we use the *round-or-cut* framework (see Carr et al. [2000], An et al. [2017], Li [2017], Chakrabarty and Negahbani [2019]): we have an exponential family of constraints with a separation oracle based on our rounding algorithm; we add a constraint from the family only if the rounding algorithm detects that it is violated. These constraints are needed to bound different moments of the cost function applied to a set of clusters generated by the first step of our rounding.

Next, we employ a slightly modified version of a reduction technique introduced by Charikar et al. [2002] which yields a simplified instance of the same problem along with a corresponding convex programming solution. After performing the reduction on the input instance and solution to the corresponding convex program, we will get a *sparsified* instance (i.e., the number of points becomes $O(k)$) along with an adjusted feasible solution that together satisfy several useful properties which relate to our convex programming relaxations and are crucial for our rounding algorithm. More details on the reduction and the properties guaranteed by it are provided in Section 3.

Finally, we perform a rounding procedure on the sparsified instance and obtain an approximate integral solution. By the properties of the reduction, using the output integral solution for the sparsified instance, we can construct an integral solution of the original instance without increasing the cost by more than a constant factor. We remark that the analysis of our rounding algorithm crucially relies on the properties guaranteed by the “non-standard” constraints we added to the relaxations of the problem in both $p \leq q$ and $p \geq q$ regimes.

1.2 Paper Organization. We start with providing the convex programming relaxations of (p, q) -Fair Clustering in Section 2. In Section 3, we concisely state the properties of the reduction by [Charikar et al., 2002] that is used in our rounding algorithm – its proof is deferred to Appendix A. Then, in Section 4, we

²This is the case, since the (ℓ_p, ℓ_∞) -cost objective is equal to the $(\ell_p, \ell_{\log n})$ -cost objective up to a constant factor.

describe our rounding algorithm and in Sections 5 and 6 we analyze the approximation guarantee of the rounding algorithm in the regimes $p \geq q$ and $p \leq q$ respectively. Finally, in Appendix B, we explore the connection of (p, q) -Fair Clustering to Min s -Union and prove polynomial hardness of approximation for the problem assuming standard hardness conjectures for Min s -Union.

2 Convex Programming Relaxations

We will use somewhat different convex programming relaxations in the two parameter regimes, when $p \geq q$ and $p \leq q$. However, the relaxations for both regimes will use the following common assignment/clustering polytope, which is often used for problems such as k -Median or Facility Location:

$\mathbf{P}_{\text{cluster}}^{\mathbf{k}} := \{(x, y) \in [0, 1]^{m \times m} \times [0, 1]^m \text{ satisfying the following}\}$

$$(2.3) \quad \sum_{\ell=1}^m x_{j\ell} = 1 \quad \forall j \in [m]$$

$$(2.4) \quad x_{j\ell} \leq y_\ell := x_{\ell\ell} \quad \forall j \in [m], \ell \in [m] \setminus \{j\}$$

$$(2.5) \quad \sum_{j=1}^m y_j \leq k$$

$$(2.6) \quad x_{j\ell} \geq 0 \quad \forall j, \ell \in [m]$$

In the intended solution, $x_{j\ell}$ is a 0-1 variable which is 1 if and only if ℓ is a center and j is assigned to a cluster centered at ℓ . The variable $y_j = x_{jj}$ is 1 if and only if j is a center.

2.1 Convex Relaxation for the Case $p \geq q$. Our algorithm will use the following convex program. We think of B as a fixed parameter of the convex program, which we can use in a binary search.

$$(2.7) \quad \begin{array}{ll} \min & B \\ \text{s.t.} & \sum_{i=1}^n z_i \leq B^q \end{array}$$

$$(2.8) \quad (x, y) \in \mathbf{P}_{\text{cluster}}^{\mathbf{k}}$$

$$(2.9) \quad z_i \geq \left(\sum_{j=1}^m w_i(j) \left(\sum_{j'=1}^m d(j, j')^q x_{jj'} \right)^{p/q} \right)^{q/p} \quad \forall i \in [n]$$

The variable z_i represents the cost incurred by group i , raised to the q . The natural constraint to express the connection between z_i and the clustering variables $x_{jj'}$ would be

$$(2.10) \quad z_i \geq \left(\sum_{j=1}^m w_i(j) \sum_{j'=1}^m d(j, j')^p x_{jj'} \right)^{q/p}.$$

However, this is not a convex constraint in the $p > q$ regime, so we further relax this connection and write Constraint (2.9) instead.

2.2 Convex Relaxation for the Case $p \leq q$. To motivate our relaxation, first consider the natural convex relaxation which is identical to the relaxation in the previous section but with Constraint (2.9) replaced by the more natural Constraint (2.10) above (which is convex in the $p \leq q$ regime). Consider even a simple case where all distances are 1, and the points can be partitioned into sets (J_1, J_2) such that $y_j = 1$ for every $j \in J_2$, and for every $j \in J_1$ there is exactly one $j' \in J_2$ such that $x_{jj'} = \varepsilon$ and $y_j = x_{jj} = 1 - \varepsilon$, and $x_{jj''} = 0$ for all $j'' \notin \{j, j'\}$. Also suppose every weight function w_i is an indicator function for some set $P_i \subseteq J_1$ of cardinality $|P_i| = t$ (see

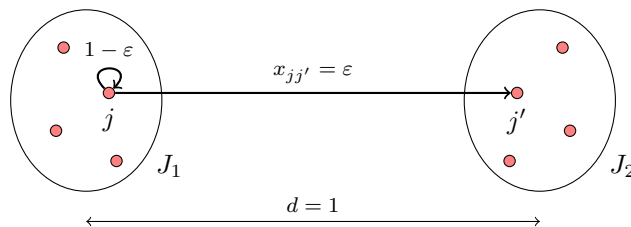


Figure 1: In this example, the natural convex program has a large integrality gap.

Figure 1). Note that in this case, the natural cost constraints would give us $z_i \geq (\sum_{j \in P_i} d(j, j')^p \cdot \varepsilon)^{q/p} = (\varepsilon \cdot t)^{q/p}$ and $B \geq \left(\sum_{i=1}^n z_i\right)^{1/q} \geq n^{1/q} \cdot (\varepsilon \cdot t)^{1/p}$. It would be natural to apply the following randomized rounding:

- assign each $j \in J_1$ to center j' independently with probability $y_{jj'} = \varepsilon$, and otherwise make j a center;
- make each $j \in J_2$ a center (since $y_j = 1$).

Denote the obtained set of centers by C . Let us see what the cost of group i w.r.t. C is. We have

$$\mathbb{E}[\text{cost}_p(C, w_i)^p] = \mathbb{E}\left[\sum_{j \in P_i} d(j, C)^p\right] = \varepsilon t.$$

Now, we are interested in the expectation $\mathbb{E}\left[\left(\sum_{j \in P_i} d(j, C)^p\right)^{q/p}\right]$, as our ultimate goal is to upper bound the expectation of $\text{cost}_{p,q}(C)^q$ (see formula (1.2)). Observe that if $\varepsilon \cdot t = \Omega(\log n)$, then $\sum_{j \in P_i} d(j, C)^p$ is concentrated around the mean and

$$\mathbb{E}\left[\left(\sum_{j \in P_i} d(j, C)^p\right)^{q/p}\right] \sim (\varepsilon t)^{q/p} = O(z_i),$$

as desired. However, if $\varepsilon t = o(1)$, then $\sum_{j=1}^m d(j, C)^p$ is close to a Poisson random variable with rate εt . Thus, the expectation of $\left(\sum_{j=1}^m d(j, C)^p\right)^{q/p}$ is at least $\sim \varepsilon t$, which is much greater than $(\varepsilon t)^{q/p}$ (in this regime). Thus,

$$\left(\mathbb{E}\left[\sum_{i=1}^n \left(\sum_{j=1}^m d(j, C)^p\right)^{q/p}\right]\right)^{1/q} \gtrsim n^{1/q} (\varepsilon t)^{1/q},$$

which is a factor of $\left(\frac{1}{\varepsilon \cdot t}\right)^{(q-p)/(pq)}$ larger than our bound for B .

To overcome this potentially polynomially-large gap, we need to introduce different constraints to handle different moments in the randomized rounding (we will use Latała's inequality in the analysis, which basically allows us to handle the two cases above separately). For instance, to decrease the integrality gap and improve the performance of the rounding w.r.t. the relaxation in the above example when $\varepsilon t = o(1)$, we may want to add constraints of the form

$$z_i \geq \sum_{j=1}^m (1 - y_j) w_i(j)^{q/p} d(j, [m] \setminus j)^q.$$

However, it is not enough to introduce such constraints for the original points, since we apply our randomized rounding to an instance produced by the reduction of Charikar et al. Loosely speaking, the reduction partitions the set of points $[m]$ into $O(k)$ initial clusters $\{V_\ell\}$; it is guaranteed that there is a k -clustering that for each ℓ , assigns all points in V_ℓ to the same center, and has cost at most a constant times greater than the optimum clustering. In our algorithm, for every ℓ , we will find one center that serves all points in V_ℓ using randomized rounding. For this approach to work, we need to introduce new constraints that depend on the partition $\{V_\ell\}$. The challenge is that sets $\{V_\ell\}$ returned by the reduction depend on the convex program solution, so we do not

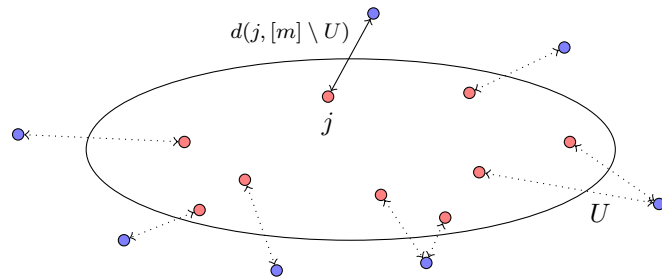


Figure 2: To compute $\text{vol}_i(U)$, we find the distance from every point j in U to the closest point outside of U .

know them when we solve the convex program. To deal with this problem, we introduce a family of exponentially-many constraints for all possible choices of sets $\{V_\ell\}$. Since we only need these constraints to hold for the set of clusters arising in the reduction and not all possible collections of clusters, we will check these constraints with a “rounding separation oracle.” That is, we will only check that the constraints hold for the set of clusters arising in the reduction, and if they do not, we will use a separating hyperplane to continue solving the convex program with the Ellipsoid Method. In order to describe our non-standard constraints, we introduce the following definition (see Figure 2).

DEFINITION 2.1. Consider a set of points $U \subset [m]$. Let $\text{vol}_i(U) = \sum_{j \in U} w_i(j) d(j, [m] \setminus U)^p$.

We briefly discuss the motivation for this definition. Consider a set of points $U \subset [m]$ (later this set will be one of the clusters returned by the reduction). Assume that our algorithm opens a set of centers C . By Definition 1.1, $\text{cost}_p(C, w_i)^p = \sum_{j=1}^m w_i(j) d(j, C)^p$. We want to lower bound the contribution of points in U to $\text{cost}_p(C, w_i)^p$ assuming that there are no centers from C in U . We show that this contribution is at least $\text{vol}_i(U)$. Indeed, observe that $d(j, C) \geq d(j, [m] \setminus U)$ for every $j \in U$, since $C \subseteq [m] \setminus U$. Therefore,

$$(2.11) \quad \sum_{j \in U} w_i(j) d(j, C)^p \geq \sum_{j \in U} w_i(j) d(j, [m] \setminus U)^p = \text{vol}_i(U).$$

In the following claim we generalize this inequality for the case where we have many disjoint sets.

CLAIM 2.1. Consider a collection $(V_\ell)_{\ell \in \Lambda}$ of pairwise disjoint subsets of $[m]$, indexed by some set of indices Λ . Let $C \subseteq [m]$ be a set of centers. Then for every group $i \in [n]$, the following lower bounds on $\text{cost}_p(C, w_i)$ hold.

$$(2.12) \quad \text{cost}_p(C, w_i)^p \geq \sum_{\ell \in \Lambda: V_\ell \cap C = \emptyset} \text{vol}_i(V_\ell) \quad \text{and}$$

$$(2.13) \quad \text{cost}_p(C, w_i)^q \geq \sum_{\ell \in \Lambda: V_\ell \cap C = \emptyset} \text{vol}_i(V_\ell)^{q/p} \quad \text{if } p \leq q.$$

Proof. Using Definition 1.1, inequality (2.11), and that sets V_ℓ are pairwise disjoint, we get,

$$\text{cost}_p(C, w_i)^p = \sum_{j=1}^m w_i(j) d(j, C)^p \geq \sum_{\ell \in \Lambda: V_\ell \cap C = \emptyset} \sum_{j \in V_\ell} w_i(j) d(j, C)^p \stackrel{\text{by (2.11)}}{\geq} \sum_{\ell \in \Lambda: V_\ell \cap C = \emptyset} \text{vol}_i(V_\ell),$$

as required. We obtain inequality (2.13) from inequality (2.12) by applying the inequality $\|a\|_1 \geq \|a\|_{q/p}$ (note that $p \leq q$) to the vector $a = (\text{vol}_i(V_\ell))_{\ell \in \Lambda: V_\ell \cap C = \emptyset}$. \square

We rewrite inequalities (2.12) and (2.13) as follows.

$$(2.14) \quad \text{cost}_p(C, w_i)^q \geq \left(\sum_{\ell \in \Lambda} \max(0, 1 - |C \cap V_\ell|) \text{vol}_i(V_\ell) \right)^{q/p},$$

$$(2.15) \quad \text{cost}_p(C, w_i)^q \geq \sum_{\ell \in \Lambda} \max(0, 1 - |C \cap V_\ell|) (\text{vol}_i(V_\ell))^{q/p}.$$

Note that the term $\max(0, 1 - |C \cap V_\ell|)$ equals 1 if $C \cap V_\ell = \emptyset$ and equals 0 otherwise. To state our convex relaxation, we will use the following notation for disjoint collections of sets:

$$\Pi(m) := \{(\Lambda, (V_\ell)_{\ell \in \Lambda}) \mid \Lambda \subseteq [m] \text{ and } (V_\ell)_{\ell \in \Lambda} \text{ are disjoint subsets of } [m]\}.$$

We are now ready to present our relaxation for the case $p \leq q$.

$$(2.16) \quad \begin{array}{ll} \min & B \\ \text{s.t.} & (x, y) \in \mathbf{P}_{\text{cluster}}^k \end{array}$$

$$(2.17) \quad \sum_{i=1}^n z_i \leq B^q$$

$$(2.18) \quad z_i \geq \left(\sum_{j=1}^m w_i(j) \sum_{j'=1}^m d(j, j')^p x_{jj'} \right)^{q/p} \quad \forall i \in [n]$$

$$(2.19) \quad z_i \geq \left(\sum_{\ell \in \Lambda} \max(0, 1 - \sum_{j \in V_\ell} y_j) \cdot \text{vol}_i(V_\ell) \right)^{q/p} \quad \forall i \in [n], \forall (\Lambda, (V_\ell)_{\ell \in \Lambda}) \in \Pi(m)$$

$$(2.20) \quad z_i \geq \sum_{\ell \in \Lambda} \max(0, 1 - \sum_{j \in V_\ell} y_j) \cdot \text{vol}_i(V_\ell)^{q/p} \quad \forall i \in [n], \forall (\Lambda, (V_\ell)_{\ell \in \Lambda}) \in \Pi(m)$$

It follows from inequalities (2.14) and (2.15) that this is a valid relaxation.

As noted in the following remark, it will be easy to check Constraints (2.19) and (2.20) once the set of clusters is fixed.

REMARK 2.1. The correctness of the reduction of Charikar et al. does not require Constraints (2.19) and (2.20), so they do not need to be checked before applying the reduction. The reduction will yield a set of centers K along with the corresponding Voronoi cells $(V_\ell)_{\ell \in K}$. To check that Constraints (2.19) and (2.20) hold for these sets for any $\Lambda \subseteq K$, it suffices to check that they hold for $\Lambda = \{\ell \in K \mid \sum_{j \in V_\ell} y_j < 1\}$.

REMARK 2.2. Let us briefly compare our relaxation with the one used by Makarychev and Vakilian for solving (p, ∞) -Fair Clustering (let us call the latter the (p, ∞) -relaxation). The (p, ∞) -relaxation has Constraints (2.16)–(2.18) and additional non-standard constraints. These non-standard constraints are essentially equivalent to Constraints (2.19) for a very special family of $(\Lambda, (V_\ell)_{\ell \in \Lambda})$: $|\Lambda| = 1$ and set V_ℓ is a ball around some point in $[m]$ (note that Constraints (2.19) and (2.20) are equivalent when $|\Lambda| = 1$). Since there are polynomially many different balls in $([m], d)$, the (p, ∞) -relaxation has polynomially-many constraints. However, if we simply adapted this relaxation to the (p, q) -Fair Clustering problem, we would get a relaxation with a polynomially large integrality gap.

3 Reduction à la Charikar et al.

We will use a slight modification of the reduction, which was used by Charikar et al. to get the first constant factor approximation algorithm for k -Median. We use the same reduction to solve the (p, q) -Fair Clustering problem in both regimes, $p \leq q$ and $p \geq q$.

We will refer to a not-necessarily-optimal solution (x, y, z) that satisfies Constraints (2.7)–(2.9) when $p \geq q$ and Constraints (2.16)–(2.18) but when $p \leq q$ as a *fractional clustering solution* (when $p \leq q$, the solution might not satisfy Constraints (2.19) and (2.20)).

THEOREM 3.1. *There is a polynomial-time reduction that given an instance \mathcal{I} of (p, q) -Fair Clustering, a fractional clustering solution (x, y, z) of value B , and a parameter $\gamma \in (0, 1/2)$ returns an instance \mathcal{I}' on a subset of points $K \subset [m]$ of size $|K| \leq \frac{k}{1-\gamma}$ with weights $w'_i(\ell)$ (where $i \in [n]$ and $\ell \in K$), and a fractional clustering solution (x', y', z') such that the following properties hold.*

1. *The cost of (x', y', z') is at most twice that of solution (x, y, z) .*

2. Let $\sigma(\ell) \in K$ be the closest point to $\ell \in K$ other than ℓ itself. The solution (x', y', z') assigns each point $\ell \in K$ only to centers ℓ and $\sigma(\ell)$. Specifically, for all $\ell \in K$:

$$x'_{\ell\ell} = y'_\ell; \quad x'_{\ell\sigma(\ell)} = 1 - y'_\ell; \quad x'_{\ell j} = 0 \text{ for } j \notin \{\ell, \sigma(\ell)\}.$$

Further, $x'_{\ell\sigma(\ell)} \leq \gamma$.

3. Let $\{V_\ell\}_{\ell \in K}$ be the Voronoi partition of $[m]$ induced by the set of centers K (Voronoi sites). Then, for every ℓ , $y'_\ell = \min\left(1, \sum_{j \in V_\ell} y_j\right)$.
4. The costs of combinatorial solutions for \mathcal{I} and \mathcal{I}' are related as follows: For every set of centers $L \subseteq [m]$ there is a set of centers $L' \subseteq K$ of cost

$$(3.21) \quad \text{cost}_{p,q}^{\mathcal{I}'}(L') \leq 2\text{cost}_{p,q}^{\mathcal{I}}(L) + \frac{4}{\gamma^{1/\nu}} B,$$

and moreover for any set of centers $L \subseteq K$, the cost of L as a solution for \mathcal{I} is bounded by

$$(3.22) \quad \text{cost}_{p,q}^{\mathcal{I}}(L) \leq \text{cost}_{p,q}^{\mathcal{I}'}(L) + \frac{2}{\gamma^{1/\nu}} B.$$

5. If $\Lambda \subseteq K$ is a set of centers such that $\sigma(\ell) \notin \Lambda$ for every $\ell \in \Lambda$, then the cost of centers $K \setminus \Lambda$ w.r.t. instance \mathcal{I} is bounded by

$$\text{cost}_{p,q}^{\mathcal{I}}(K \setminus \Lambda) \leq \frac{6}{\gamma^{1/\nu}} B + 2 \left(\sum_{i=1}^n \left(\sum_{\ell \in \Lambda} \text{vol}_i(V_\ell) \right)^{q/p} \right)^{1/q}.$$

Charikar et al. presented this reduction and proved that it satisfies properties 1–3 and (3.22). As far as we know, Property (3.21) has not been explicitly stated or used before, but its proof is similar to that of Property (3.22). Finally, Property 5 is new – we need it to use our new CP Constraints (2.19) and (2.20). We will also use the following observation from [Charikar et al., 2002].

OBSERVATION 3.1. We can efficiently find a partition (K_1, K_2) of K such that $\sigma(K_1) \subseteq K_2$, $\sigma(K_2) \subseteq K_1$, and

$$\sum_{\ell \in K_1} x'_{\ell\sigma(\ell)} \geq \frac{|K| - k}{2}.$$

For completeness, we provide proofs of Theorem 3.1 and Observation 3.1 in Appendix A.

4 A Randomized Rounding

In this section, we describe two steps that will be used in the rounding algorithms for both parameter regimes. All our algorithms will begin by applying the reduction from Theorem 3.1 for $\gamma = 1/5$. Let K be the set of points obtained by applying this reduction, let (x', y', z') be the corresponding convex programming solution, as described in the theorem, and let K_1 be as in Observation 3.1. Note that if $|K| \leq k$, then the current centers already give a constant factor approximation (since their cost in the new instance is 0, and so by Property (3.22) in Theorem 3.1, the cost in the original instance is at most $2B/\gamma^{1/\nu} \leq 10B$ (and we can simply add $|K| - k$ centers to K from $[m] \setminus K$ in the original instance only reducing the cost). Thus, we assume that $|K| > k$ below. Consider the following rounding:

- Let

$$K' = \left\{ \ell \in K_1 \mid x'_{\ell\sigma(\ell)} \geq \frac{|K| - k}{4|K_1|} \right\},$$

and note by Observation 3.1 that

$$\sum_{\ell \in K'} x'_{\ell\sigma(\ell)} \geq \sum_{\ell \in K_1} x'_{\ell\sigma(\ell)} - \frac{|K| - k}{4|K_1|} \cdot |K_1 \setminus K'| \geq \frac{|K| - k}{2} - \frac{|K| - k}{4} \geq \frac{|K| - k}{4}.$$

- Let $L = K'$. Independently close (remove from L) every center $\ell \in K'$ with probability $5x'_{\ell\sigma(\ell)}$ (recall that $x'_{\ell\sigma(\ell)} \leq \gamma = 1/5$ by item 2 of Theorem 3.1).
- If $|L| < k$, reopen (add to L) an arbitrary collection of $k - |L|$ centers at no additional cost.

The expected number of centers that we close in the second step is $5 \sum_{\ell \in K'} x'_{\ell\sigma(\ell)} \geq \frac{5}{4}(|K| - k)$. By a Chernoff bound this number is at least $|K| - k$ centers w.h.p. (unless $|K| - k$ is bounded by some sufficiently large constant, in which case we can enumerate over all possible solutions to get a constant factor approximation). While the analysis of the cost of this rounding depends on the specific parameter regime we are in, we introduce the following notation which we will use in both cases. For every group i and center $\ell \in K'$, we define random variables specifying the per-center and total costs incurred to every group $i \in [n]$ by this rounding:

$$(4.23) \quad Z_{i\ell} := \begin{cases} \text{vol}_i(V_\ell) & \text{if we close center } \ell, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad Z_i = \sum_{\ell \in K'} Z_{i\ell}.$$

For the simpler analysis in Section 5, it will suffice to analyze the cost in the new instance \mathcal{I}' (produced by the reduction from Theorem 3.1), for which we define the following variables:

$$(4.24) \quad Z'_{i\ell} := \begin{cases} w'_i(\ell)d(\ell, \sigma(\ell))^p & \text{if we close center } \ell, \\ 0 & \text{otherwise,} \end{cases} \quad \text{and} \quad Z'_i = \sum_{\ell \in K'} Z'_{i\ell}.$$

CLAIM 4.1. *The (ℓ_p, ℓ_q) -cost in \mathcal{I} of the clustering found by randomized rounding is bounded by³*

$$\min \left\{ 30B + 2 \left(\sum_{i=1}^n Z_i^{q/p} \right)^{1/q}, 10B + \left(\sum_{i=1}^n (Z'_i)^{q/p} \right)^{1/q} \right\}$$

Proof. Note that if we close a center ℓ then we do not close $\sigma(\ell)$, since (i) we only close centers ℓ in $K' \subseteq K_1$ and (ii) if $\ell \in K'$ then $\sigma(\ell) \in K_2$. Therefore, we can apply item 5 of Theorem 3.1 with $\Lambda = K \setminus L$ (the set of centers we closed). We get,

$$\text{cost}_{p,q}^{\mathcal{I}}(L) \leq \frac{6}{\gamma^{1/\nu}} B + 2 \left(\sum_{i=1}^n \left(\sum_{\ell \notin L} \text{vol}_i(V_\ell) \right)^{q/p} \right)^{1/q} = \frac{6}{\gamma^{1/\nu}} B + 2 \left(\sum_{i=1}^n Z_i^{q/p} \right)^{1/q}.$$

where $\nu = \min(p, q)$, $\gamma = 1/5$, and thus $6/\gamma^{1/\nu} \leq 30$. Also, the cost of L w.r.t. instance \mathcal{I}' is

$$\left(\sum_{i=1}^n \left(\sum_{\ell \in K} w'_i(\ell)d(\ell, L)^p \right)^{q/p} \right)^{1/q} = \left(\sum_{i=1}^n \left(\sum_{\ell \in K \setminus L} w'_i(\ell)d(\ell, \sigma(\ell))^p \right)^{q/p} \right)^{1/q} = \left(\sum_{i=1}^n (Z'_i)^{q/p} \right)^{1/q}.$$

Here, we used that $d(\ell, L) = 0$ if ℓ is not closed and $d(\ell, L) = d(\ell, \sigma(\ell))$, otherwise. By Theorem 3.1, item 4, the cost of L w.r.t. instance \mathcal{I} is

$$\text{cost}_{p,q}^{\mathcal{I}}(L) \leq \frac{2}{\gamma^{1/\nu}} B + \left(\sum_{i=1}^n (Z'_i)^{q/p} \right)^{1/q} \leq 10B + \left(\sum_{i=1}^n (Z'_i)^{q/p} \right)^{1/q}.$$

□

³Note that terms $30B$ and $10B$ only add a constant to the approximation factor of the randomized rounding scheme, since B is at most the cost of the optimal clustering. Thus, the main challenge will be to upper bound either $\left(\sum_{i=1}^n Z_i^{q/p} \right)^{1/q}$ or $\left(\sum_{i=1}^n (Z'_i)^{q/p} \right)^{1/q}$.

5 Approximation Algorithm for the Case $p \geq q$

In this section, we present our rounding algorithm for the convex program from Section 2.1.

We start out by applying the reduction described in Theorem 3.1. We now consider two cases. If $|K| - k \geq \sqrt{k}$, then we apply the randomized rounding from Section 4. Otherwise, we re-weight the points and run an approximation algorithm for k -Clustering with ℓ_q -norm objective. Let us first analyze the performance of randomized rounding when $p \geq q$.

LEMMA 5.1. *The expected cost of the randomized rounding is at most $O\left(B \cdot (k/(|K| - k))^{\frac{p-q}{pq}}\right)$.*

Proof. By Claim 4.1, it suffices to bound the expectation of $(\sum_{i=1}^n (Z'_i)^{q/p})^{1/q}$. Let us do that now. By Jensen's inequality, and linearity of expectation, we have

$$(5.25) \quad \mathbb{E} \left[\left(\sum_{i=1}^n (Z'_i)^{q/p} \right)^{1/q} \right] \leq \left(\mathbb{E} \left[\sum_{i=1}^n (Z'_i)^{q/p} \right] \right)^{1/q} = \left(\sum_{i=1}^n \mathbb{E} \left[(Z'_i)^{q/p} \right] \right)^{1/q} \leq \left(\sum_{i=1}^n \mathbb{E} [Z'_i]^{q/p} \right)^{1/q}$$

Let us now bound the expectation $\mathbb{E}[Z'_i]$. First, recall that by definition of K' and Theorem 3.1, for every $\ell \in K'$, we have

$$(5.26) \quad x'_{\ell\sigma(\ell)} \geq \frac{|K| - k}{4|K_1|} \geq \frac{|K| - k}{4|K|} \geq \frac{|K| - k}{4k/(1-\gamma)} = \frac{|K| - k}{5k}.$$

From the definition of Z'_i , using that (x', y', z') satisfies Constraint (2.9) by Theorem 3.1, we have

$$\begin{aligned} \mathbb{E}[Z'_i] &= \sum_{\ell \in K'} \mathbb{E}[Z'_{i\ell}] = 5 \sum_{\ell \in K'} x'_{\ell\sigma(\ell)} w'_i(\ell) d(\ell, \sigma(\ell))^p \\ &\leq 5 \left(\frac{5k}{|K| - k} \right)^{\frac{p-q}{q}} \sum_{\ell \in K'} (x'_{\ell\sigma(\ell)})^{p/q} w'_i(\ell) d(\ell, \sigma(\ell))^p \quad \text{by (5.26)} \\ &= 5 \left(\frac{5k}{|K| - k} \right)^{\frac{p-q}{q}} (z'_i)^{p/q}. \quad \text{by Constraint (2.9)} \end{aligned}$$

Plugging this bound back into (5.25), we get

$$\mathbb{E} \left[\left(\sum_{i=1}^n (Z'_i)^{q/p} \right)^{1/q} \right] \leq \left(5^{q/p} \left(\frac{5k}{|K| - k} \right)^{\frac{p-q}{p}} z'_i \right)^{1/q} = 5^{1/p} \left(\frac{5k}{|K| - k} \right)^{\frac{p-q}{pq}} \cdot \left(\sum_{i=1}^n z'_i \right)^{1/q},$$

By Constraint (2.7) and Theorem 3.1 $(\sum_{i=1}^n z'_i)^{1/q} \leq 2B$. We get

$$\mathbb{E} \left[\left(\sum_{i=1}^n (Z'_i)^{q/p} \right)^{1/q} \right] \leq 10 \left(\frac{5k}{|K| - k} \right)^{\frac{p-q}{pq}} \cdot B,$$

as required. \square

Thus, as mentioned earlier, the randomized rounding indeed gives the desired approximation when $|K| - k \geq \sqrt{k}$. Let us see a different rounding algorithm, which gives the desired guarantee when $|K| - k \leq \sqrt{k}$. In this rounding, we define a new weight function $\hat{w} : K \rightarrow \mathbb{R}_{\geq 0}$ as follows:

$$\hat{w}(\ell) := \sum_{i=1}^n w'_i(\ell)^{q/p}.$$

Recall that for any $q \in [1, \infty)$, k -Clustering with ℓ_q -cost can be approximated up to a constant factor (see Remark 1.1). Our rounding algorithm in this case is simple:

- Apply a constant-factor approximation for k -Clustering with ℓ_q -cost to the current input (on K) with new weights \hat{w} and return the set of centers L chosen by this algorithm.

Let us analyze the approximation guarantee. Recall that the new instance (obtained by the reduction from Theorem 3.1) has optimum value $B' = O(\gamma^{-1/q})(B^* + B) = O(B^*)$, where B^* is the optimum value of the original instance. Then for the above algorithm applied to the new instance, we have the following guarantee.

LEMMA 5.2. *The cost $\text{cost}_{p,q}^{\mathcal{I}'}(L)$ is at most $O\left(B' \cdot (|K| - k)^{\frac{p-q}{pq}}\right)$.*

Proof. Since we use a constant-factor approximation for the ℓ_q objective, it suffices to show that for every set $\hat{K} \subset K$ of k centers, the following values

- the (ℓ_p, ℓ_q) -cost of \hat{K} w.r.t. the original weights w'_i (from the reduction of Theorem 3.1), and
- the ℓ_q -cost of \hat{K} w.r.t. the new weights \hat{w}_i

are within a factor of $(|K| - k)^{\frac{p-q}{pq}}$ of each other. Indeed, let $\hat{K} \subseteq K$ be any set of k centers. Then

$$\begin{aligned} \text{cost}_{p,q}^{\mathcal{I}'}(\hat{K}) &= \left(\sum_{i=1}^n \left(\sum_{\ell \in K} w'_i(\ell) d(\ell, \hat{K})^p \right)^{q/p} \right)^{1/q} \\ &\leq \left(\sum_{i=1}^n \sum_{\ell \in K} w'_i(\ell)^{q/p} d(\ell, \hat{K})^q \right)^{1/q} && \text{since } \|\cdot\|_{p/q} \leq \|\cdot\|_1 \\ &= \left(\sum_{\ell \in K} \underbrace{\left(\sum_{i=1}^n w'_i(\ell)^{q/p} \right)}_{\hat{w}(\ell)} d(\ell, \hat{K})^q \right)^{1/q}, \end{aligned}$$

which is exactly the ℓ_q objective of our new instance applied to this set of centers.

On the other hand, again for any set $\hat{K} \subseteq K$ of k centers, we can bound the q -norm objective of the instance with weights \hat{w} as follows

$$\begin{aligned} \left(\sum_{\ell \in K} \hat{w}(\ell) d(\ell, \hat{K})^q \right)^{\frac{1}{q}} &= \left(\sum_{\ell \in K \setminus \hat{K}} \hat{w}(\ell) d(\ell, \hat{K})^q \right)^{\frac{1}{q}} = \left(\sum_{\ell \in K \setminus \hat{K}} \sum_{i=1}^n w'_i(\ell)^{q/p} d(\ell, \hat{K})^q \right)^{\frac{1}{q}} \\ &= \left(\sum_{i=1}^n \sum_{\ell \in K \setminus \hat{K}} w'_i(\ell)^{q/p} d(\ell, \hat{K})^q \right)^{\frac{1}{q}} \\ &\stackrel{\text{by Hölder}}{\leq} \left(\sum_{i=1}^n \left(\sum_{\ell \in K \setminus \hat{K}} 1^{\frac{p}{p-q}} \right)^{\frac{p-q}{p}} \left(\sum_{\ell \in K \setminus \hat{K}} w'_i(\ell) d(\ell, \hat{K})^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} \\ &= |K \setminus \hat{K}|^{\frac{p-q}{pq}} \left(\sum_{i=1}^n \left(\sum_{\ell \in K} w'_i(\ell) d(\ell, \hat{K})^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}} = (|K| - k)^{\frac{p-q}{pq}} \text{cost}_{p,q}^{\mathcal{I}'}(\hat{K}), \end{aligned}$$

and the proof follows. \square

Proof. (Proof of Theorem 1.1.) We solve the convex programming relaxation for the problem and then apply the reduction from Theorem 3.1. If the set of points K found by the reduction is of cardinality $|K| \leq k$, we add $k - |K|$ points from $[m]$ and return the resulting set, which as noted earlier has cost $O(B)$. Otherwise, if $|K| - k > \sqrt{k}$, we run the randomized rounding procedure, which by Lemma 5.1 yields an $O(k^{(p-q)/(2pq)})$ -approximation. Finally,

if $0 < |K| - k \leq \sqrt{k}$, we run a constant-factor approximation for the ℓ_q norm k -Clustering with weights \hat{w} and obtain a solution L . Let $B_{\mathcal{I}'}^*$ be the (ℓ_p, ℓ_q) -cost of the optimal solution for \mathcal{I}' and B^* be the (ℓ_p, ℓ_q) -cost of the optimal solution for \mathcal{I} . By Lemma 5.2,

$$\text{cost}_{p,q}^{\mathcal{I}'}(L) \leq O(k^{(p-q)/(2pq)})B_{\mathcal{I}'}^*.$$

By Theorem 3.1, item 4:

$$\begin{aligned} \text{cost}_{p,q}(L) &\leq \text{cost}_{p,q}^{\mathcal{I}'}(L) + 10B \\ &\stackrel{(3.22)}{\leq} O(k^{(p-q)/(2pq)})B_{\mathcal{I}'}^* + 10B \\ &\stackrel{(3.21)}{\leq} O(k^{(p-q)/(2pq)})(2B^* + 20B) + 10B = O(k^{(p-q)/(2pq)})B^*, \end{aligned}$$

here we used that $B \leq B^*$. We conclude that the algorithm gives an $O(k^{\frac{p-q}{pq}})$ approximation, as required. \square

6 Approximation Algorithm for the Case $p \leq q$

In this section, we upper bound the cost of the solution produced by the rounding procedure. In the analysis, we will use Latała's inequality.

THEOREM 6.1. (LATAŁA [1997], COROLLARY 3) *There exists a universal constant M such that if Z_1, \dots, Z_N are independent non-negative random variables and $\alpha \geq 1$, then*

$$\left(\mathbb{E} \left[\left(\sum_{i=1}^N Z_i \right)^\alpha \right] \right)^{1/\alpha} \leq \frac{M\alpha}{\ln(1+\alpha)} \max \left(\sum_{i=1}^N \mathbb{E}[Z_i], \left(\sum_{i=1}^N \mathbb{E}[Z_i^\alpha] \right)^{1/\alpha} \right).$$

We will use Latała's inequality with $\alpha = q/p$. Denote $M_{pq} = \left(\frac{Mq}{p \ln(1+q/p)} \right)^{q/p}$. Then,

$$(6.27) \quad \mathbb{E} \left[\left(\sum_{i=1}^N Z_i \right)^{q/p} \right] \leq M_{pq} \left(\left(\sum_{i=1}^N \mathbb{E}[Z_i] \right)^{q/p} + \sum_{i=1}^N \mathbb{E}[Z_i^{q/p}] \right).$$

The algorithm for (p, q) -Fair Clustering in the case $p \leq q$ simply solves the convex problem, applies the reduction from Theorem 3.1, and then runs the randomized rounding procedure. We denote the obtained set of centers by L . Now we are ready to prove the main result of this section.

LEMMA 6.1. *The rounding procedure outputs a solution for \mathcal{I} of cost at most $O \left(\left(\frac{q}{\ln(1+q/p)} \right)^{1/p} \right) B$ in expectation.*

Proof. We use random variables $Z_{i\ell}$ defined in (4.23). By Claim 4.1, the cost of the solution for instance \mathcal{I}' found by the rounding procedure is $\left(\sum_{i=1}^n \left(\sum_{\ell \in K'} Z_{i\ell} \right)^{q/p} \right)^{1/q}$. We upper bound this cost using Latała's inequality (6.27). For every $i \in [m]$, we have

$$\mathbb{E} \left[\left(\sum_{\ell \in K'} Z_{i\ell} \right)^{q/p} \right] \leq M_{pq} \left(\left(\mathbb{E} \left[\sum_{\ell \in K'} Z_{i\ell} \right] \right)^{q/p} + \sum_{\ell \in K'} \mathbb{E}[Z_{i\ell}^{q/p}] \right)$$

Now,

$$\mathbb{E} \left[\sum_{i=1}^n \left(\sum_{\ell \in K'} Z_{i\ell} \right)^{q/p} \right] \leq M_{pq} \left(\sum_{i=1}^n \left(\mathbb{E} \left[\sum_{\ell \in K'} Z_{i\ell} \right] \right)^{q/p} + \sum_{i=1}^n \sum_{\ell \in K'} \mathbb{E}[Z_{i\ell}^{q/p}] \right)$$

Using that $\mathbb{E}[Z_{i\ell}] = 5x'_{\ell\sigma(\ell)} \cdot \text{vol}(V_\ell)$, we get

$$\sum_{i=1}^n \left(\mathbb{E} \left[\sum_{\ell \in K'} Z_{i\ell} \right] \right)^{q/p} \leq 5^{q/p} \sum_{i=1}^n \left(\sum_{\ell \in K'} x'_{\ell\sigma(\ell)} \cdot \text{vol}_i(V_\ell) \right)^{q/p}.$$

By items 2 and 3 of Theorem 3.1, $x'_{\ell\sigma(\ell)} = 1 - y'_\ell = \max(0, 1 - \sum_{j \in V_\ell} y_j)$, so from Constraint (2.19) we get

$$\sum_{i=1}^n \left(\mathbb{E} \left[\sum_{\ell \in K'} Z_{i\ell} \right] \right)^{q/p} \leq 5^{q/p} \sum_{i=1}^n z_i \leq 5^{q/p} B^q.$$

We also have,

$$\sum_{i=1}^n \sum_{\ell \in K'} \mathbb{E}[Z_{i\ell}^{q/p}] \leq 5^{q/p} \sum_{i=1}^n \sum_{\ell \in K'} x'_{\ell\sigma(\ell)} \cdot \text{vol}_i(V_\ell)^{q/p}.$$

And by the same argument as above, but using Constraint (2.20), we also get

$$\sum_{i=1}^n \sum_{\ell \in K'} \mathbb{E}[Z_{i\ell}^{q/p}] \leq 5^{q/p} \sum_{i=1}^n z_i \leq 5^{q/p} B^q.$$

We conclude that the expected cost of the clustering is at most,

$$\mathbb{E} \left[\left(\sum_{i=1}^n \left(\sum_{\ell \in K'} Z_{i\ell} \right)^{q/p} \right)^{1/q} \right] \stackrel{\text{Jensen's inequality}}{\leq} \left(\mathbb{E} \left[\sum_{i=1}^n \left(\sum_{\ell \in K'} Z_{i\ell} \right)^{q/p} \right] \right)^{1/q} \leq O(M_{pq}^{1/q}) B = O\left(\left(\frac{q}{\ln(1+q/p)}\right)^{1/p}\right) B.$$

□

Proof. (Proof of Theorem 1.2.) We solve the convex programming relaxation for the problem, apply the reduction, and run the randomized rounding procedure. This gives us a solution of cost at most $O\left(\left(\frac{q}{\ln(1+q/p)}\right)^{1/p}\right) B$ in expectation. □

Acknowledgments

The authors would like to thank the anonymous reviewers for their helpful suggestions on improving the notation and presentation of these results.

References

- M. Abbasi, A. Bhaskara, and S. Venkatasubramanian. Fair clustering via equitable group representations. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 504–514, 2021.
- S. Ahmadian, A. Epasto, R. Kumar, and M. Mahdian. Clustering without over-representation. In *Proceedings of the SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 267–275, 2019.
- S. Ahmadian, A. Norouzi-Fard, O. Svensson, and J. Ward. Better guarantees for k -means and euclidean k -median by primal-dual algorithms. *SIAM Journal on Computing*, 49(4), 2020.
- H.-C. An, M. Singh, and O. Svensson. LP-based algorithms for capacitated facility location. *SIAM Journal on Computing*, 46(1):272–306, 2017.
- B. Anthony, V. Goyal, A. Gupta, and V. Nagarajan. A plant location guide for the unsure: Approximation algorithms for min-max location problems. *Mathematics of Operations Research*, 35(1):79–101, 2010.
- A. Backurs, P. Indyk, K. Onak, B. Schieber, A. Vakilian, and T. Wagner. Scalable fair clustering. In *Proceedings of the International Conference on Machine Learning*, pages 405–413, 2019.
- S. Bera, D. Chakrabarty, N. Flores, and M. Negahbani. Fair algorithms for clustering. In *Advances in Neural Information Processing Systems*, pages 4955–4966, 2019.
- I. O. Bercea, M. Groß, S. Khuller, A. Kumar, C. Rösner, D. R. Schmidt, and M. Schmidt. On the cost of essentially fair clusterings. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, 2019.

- A. Bhaskara, M. Charikar, V. Guruswami, A. Vijayaraghavan, and Y. Zhou. Polynomial integrality gaps for strong sdp relaxations of densest k -subgraph. In *Proceedings of the Symposium on Discrete Algorithms*, pages 388–405, 2012.
- S. Bhattacharya, P. Chalmersook, K. Mehlhorn, and A. Neumann. New approximability results for the robust k -median problem. In *Scandinavian Workshop on Algorithm Theory*, pages 50–61, 2014.
- B. Brubach, D. Chakrabarti, J. Dickerson, S. Khuller, A. Srinivasan, and L. Tsepenekas. A pairwise fair and community-preserving approach to k -center clustering. In *Proceedings of the International Conference on Machine Learning*, pages 1178–1189, 2020.
- R. D. Carr, L. K. Fleischer, V. J. Leung, and C. A. Phillips. Strengthening integrality gaps for capacitated network design and covering problems. In *Proceedings of the Symposium on Discrete algorithms*, pages 106–115, 2000.
- D. Chakrabarty and M. Negahbani. Generalized center problems with outliers. *ACM Transactions on Algorithms (TALG)*, 15(3):1–14, 2019.
- D. Chakrabarty and C. Swamy. Approximation algorithms for minimum norm and ordered optimization problems. In *Proceedings of the Symposium on Theory of Computing*, pages 126–137, 2019.
- M. Charikar, S. Guha, É. Tardos, and D. B. Shmoys. A constant-factor approximation algorithm for the k -median problem. *Journal of Computer and System Sciences*, 65(1):129–149, 2002.
- X. Chen, B. Fain, L. Lyu, and K. Munagala. Proportionally fair clustering. In *Proceedings of the International Conference on Machine Learning*, pages 1032–1041, 2019.
- F. Chierichetti, R. Kumar, S. Lattanzi, and S. Vassilvitskii. Fair clustering through fairlets. In *Advances in Neural Information Processing Systems*, pages 5036–5044, 2017.
- E. Chlamtác, M. Dinitz, and Y. Makarychev. Minimizing the union: Tight approximations for small set bipartite vertex expansion. In *Proceedings of the Symposium on Discrete Algorithms*, pages 881–899, 2017.
- E. Chlamtác, M. Dinitz, C. Konrad, G. Kortsarz, and G. Rabanca. The densest k -subhypergraph problem. *SIAM Journal on Discrete Mathematics*, 32(2):1458–1477, 2018.
- E. Chlamtác and P. Manurangsi. Sherali-Adams Integrality Gaps Matching the Log-Density Threshold. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 116, pages 10:1–10:19, 2018.
- M. Ghadiri, S. Samadi, and S. Vempala. Socially fair k -means clustering. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 438–448, 2021.
- T. F. Gonzalez. Clustering to minimize the maximum intercluster distance. *Theoretical computer science*, 38: 293–306, 1985.
- D. Goyal and R. Jaiswal. FPT approximation for socially fair clustering. *arXiv preprint arXiv:2106.06755*, 2021.
- A. Gupta and K. Tangwongsan. Simpler analyses of local search algorithms for facility location. *arXiv preprint arXiv:0809.2554*, 2008.
- D. S. Hochbaum and D. B. Shmoys. A best possible heuristic for the k -center problem. *Mathematics of operations research*, 10(2):180–184, 1985.
- L. Huang, S. Jiang, and N. Vishnoi. Coresets for clustering with fairness constraints. In *Proceedings of the Conference on Neural Information Processing Systems*, 2019.
- C. Jung, S. Kannan, and N. Lutz. A center in your neighborhood: Fairness in facility location. In *Proceedings of the Symposium on Foundations of Responsible Computing*, page 5:1–5:15, 2020.

- T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu. A local search approximation algorithm for k -means clustering. *Computational Geometry*, 28(2-3):89–112, 2004.
- M. Kleindessner, P. Awasthi, and J. Morgenstern. Fair k -center clustering for data summarization. In *Proceedings of the International Conference on Machine Learning*, pages 3448–3457, 2019.
- M. Kleindessner, P. Awasthi, and J. Morgenstern. A notion of individual fairness for clustering. *arXiv preprint arXiv:2006.04960*, 2020.
- R. Latała. Estimation of moments of sums of independent real random variables. *The Annals of Probability*, 25(3):1502–1513, 1997.
- S. Li. On uniform capacitated k -median beyond the natural LP relaxation. *ACM Transactions on Algorithms (TALG)*, 13(2):1–18, 2017.
- S. Li and O. Svensson. Approximating k -median via pseudo-approximation. *SIAM Journal on Computing*, 45(2):530–547, 2016.
- S. Mahabadi and A. Vakilian. Individual fairness for k -clustering. In *Proceedings of the International Conference on Machine Learning*, pages 6586–6596, 2020.
- Y. Makarychev and A. Vakilian. Approximation algorithms for socially fair clustering. In *Proceedings of the Conference on Learning Theory*, pages 3246–3264. PMLR, 2021.
- E. Micha and N. Shah. Proportionally fair clustering revisited. In *Proceedings of the International Colloquium on Automata, Languages, and Programming*, pages 85:1–85:16, 2020.
- M. Schmidt, C. Schwiegelshohn, and C. Sohler. Fair coresets and streaming algorithms for fair k -means. In *Proceedings of the International Workshop on Approximation and Online Algorithms*, pages 232–251, 2019.

A Proof of Theorem 3.1 and Observation 3.1

Proof. (Proof of Theorem 3.1.) Let $\nu = \min(p, q)$. We define the Convex Program (CP) cost of point $j \in [m]$ as

$$(A.1) \quad C(j, x) \equiv C(j) = \left(\sum_{j' \in [m]} x_{jj'} d(j, j')^\nu \right)^{1/\nu}.$$

We sort all points according to the value of $C(j)$. Renaming the points if necessary, we may assume that

$$C(1) \leq \dots \leq C(m).$$

Now we choose a subset of points K and assign each point to exactly one point in K . Initially, $K = \emptyset$ and all points are unassigned. Then we process points one by one, starting with 1 and ending with m . When we process point j , we perform the following steps if j has not been assigned to any vertex j' yet.

- Add j to K and assign j to j .
- For every unassigned $j' > j$, if $d(j, j') \leq \frac{2}{\gamma^{1/\nu}} C(j')$, assign j' to j .

After all the points are processed, each of them is assigned to some $\ell \in K$ (some points are assigned to themselves). For $\ell \in K$, let

- U_ℓ be the set of points assigned to ℓ .
- V_ℓ be the set of points that are closer to ℓ than to any other point in K (we break ties arbitrarily).
- $\sigma(\ell)$ be the closest point to ℓ in K other than ℓ itself. We break ties arbitrarily but consistently; then the set of edges $(\ell, \sigma(\ell))$ forms a forest.

For a set of points $A \subseteq [m]$, denote $w_i(A) = \sum_{j \in A} w_i(j)$. Define new weights w'_ℓ for $\ell \in K$ by

$$w'_i(\ell) = w_i(U_\ell) = \sum_{j \in U_\ell} w_i(j).$$

We obtained the desired instance \mathcal{I}' of (p, q) -Fair Clustering on K with weights $\{w'_i(j)\}_{j \in K}$. Now we define the CP solution.

$$\begin{aligned} y'_\ell &= \min(1, \sum_{j \in V_\ell} y_j) \\ x'_{\ell\ell} &= y'_\ell; \quad x'_{\ell\sigma(\ell)} = 1 - y'_\ell; \quad x'_{\ell j} = 0 \quad \text{for } j \notin \{\ell, \sigma(\ell)\} \\ z'_i &= 2^q z_i \end{aligned}$$

We verify that all the required properties hold.

We verify that (x', y', z') is a fractional clustering solution and item 1 holds. From the definition of (x', y', z') , it follows right away that Constraints (2.3), (2.4), (2.5), and (2.6) are satisfied. We need to check that Constraints (2.9) and (2.18). We rewrite CP Constraints (2.9) and (2.18) uniformly using notation $C(j)$.

$$(A.2) \quad z_i \geq \left(\sum_j w_i(j) C(j)^p \right)^{q/p}.$$

Note that for all $j \in U_\ell$, $j \geq \ell$ and therefore $C(\ell) \leq C(j)$. We have,

$$z_i^{p/q} \geq \sum_{j \in [m]} w_i(j) C(j)^p = \sum_{\ell \in K} \sum_{j \in U_\ell} w_i(j) C(j)^p \geq \sum_{\ell \in K} \sum_{j \in U_\ell} w_i(j) C(\ell)^p = \sum_{\ell \in K} w'_i(\ell) C(\ell)^p.$$

Now consider $\ell \in K$ and $j \notin V_\ell$. Let us say $j \in V_{\ell'}$. Then $d(\ell, j) \geq d(\ell', j)$ and $d(\ell, j) + d(\ell', j) \geq d(\ell, \ell')$. Therefore,

$$(A.3) \quad d(\ell, j) \geq d(\ell, \ell')/2 \geq d(\ell, \sigma(\ell))/2$$

here we used that $\sigma(\ell)$ is the closest to ℓ point in K other than ℓ itself. We have,

$$C(\ell)^\nu = \sum_{j \in [m]} d(\ell, j)^\nu x_{\ell j} \geq \sum_{j \notin V_\ell} d(\ell, j)^\nu x_{\ell j} \geq \sum_{j \notin V_\ell} \left(\frac{d(\ell, \sigma(\ell))}{2} \right)^\nu x_{\ell j} = \left(\frac{d(\ell, \sigma(\ell))}{2} \right)^\nu \sum_{j \notin V_\ell} x_{\ell j}.$$

If $y'_\ell < 1$ then

$$\sum_{j \notin V_\ell} x_{\ell j} = 1 - \sum_{j \in V_\ell} x_{\ell j} \geq 1 - \sum_{j \in V_\ell} y_j = 1 - y'_\ell = x'_{\ell\sigma(\ell)}.$$

If $y'_\ell = 1$, then $\sum_{j \notin V_\ell} x_{\ell j} \geq 0 = 1 - y'_\ell = x'_{\ell\sigma(\ell)}$. In either case,

$$(A.4) \quad \sum_{j \notin V_\ell} x_{\ell j} \geq 1 - y'_\ell = x'_{\ell\sigma(\ell)}.$$

Similarly to (A.1), define $C'(\ell, x') \equiv C'(\ell) = (\sum_{\ell' \in K} x'_{\ell\ell'} d(\ell, \ell')^\nu)^{1/\nu} = d(\ell, \sigma(\ell)) \cdot (x'_{\ell\sigma(\ell)})^{1/\nu}$. Then,

$$C(\ell)^\nu \geq \frac{1}{2^\nu} d(\ell, \sigma(\ell))^\nu x'_{\ell\sigma(\ell)} = \frac{1}{2^\nu} C'(\ell)^\nu.$$

We conclude that $C(\ell) \geq C'(\ell)/2$ and hence $z_i^{p/q} \geq \frac{1}{2^p} \sum_{\ell \in K} w'_i(\ell) C'(\ell)^p$. Therefore,

$$(z'_i)^{p/q} = 2^p \cdot z_i^{p/q} \geq 2^p \cdot \frac{1}{2^p} \sum_{\ell \in K} w'_i(\ell) C'(\ell)^p = \sum_{\ell \in K} w'_i(\ell) C'(\ell)^p,$$

as required by CP Constraints (2.9) and (2.18). Note that since $z'_i = 2^q z_i$, cost B' of solution (x', y', z') is at most twice that of (x, y, z) .

Now we verify that item 2 holds and $|K| \leq k/(1 - \gamma)$. Formulas for x' and y' follow from their definitions. First, we show that $1 - y'_\ell \leq \gamma$. By (A.4), $1 - y'_\ell \leq \sum_{j \notin V_\ell} x_{\ell j}$. Applying (A.3), we get

$$1 - y'_\ell \leq \sum_{j \notin V_\ell} x_{\ell j} \leq \frac{\sum_{j \notin V_\ell} x_{\ell j} d(\ell, j)^\nu}{(d(\ell, \sigma(\ell))/2)^\nu} \leq \frac{(2C(\ell))^\nu}{d(\ell, \sigma(\ell))^\nu}.$$

Now, both points ℓ and $\sigma(\ell)$ are in K . Therefore, neither ℓ was assigned to $\sigma(\ell)$ nor $\sigma(\ell)$ was assigned to ℓ . This means that $d(\ell, \sigma(\ell)) > \frac{2}{\gamma^{1/\nu}} \max(C(\ell), C(\sigma(\ell))) \geq \frac{2C(\ell)}{\gamma^{1/\nu}}$. We conclude that $1 - y'_\ell \leq \gamma$, as required. Finally, the bound $|K| \leq k/(1 - \gamma)$ follows from the following inequality $k \geq \sum_{\ell \in K} y'_\ell \geq (1 - \gamma)|K|$.

It is immediate that item 3 holds.

Now we verify that item 4 holds. We first prove inequality (3.21). Consider a set of centers $L \subset [m]$. Note that points in L are not necessarily in K . So L is not necessarily a valid set of centers for instance \mathcal{I}' . However, we can think of L as a set of Steiner centers and then compute the cost of L with respect to instance \mathcal{I}' . Formally, we write

$$\begin{aligned} \left(\sum_{i=1}^n \left(\sum_{\ell \in K} w'_i(\ell) \cdot d(\ell, L)^p \right)^{q/p} \right)^{1/q} &= \left(\sum_{i=1}^n \left(\sum_{\ell \in K} \sum_{j \in U_\ell} w_i(j) \cdot d(\ell, L)^p \right)^{q/p} \right)^{1/q} \\ &\leq \left(\sum_{i=1}^n \left(\sum_{\ell \in K} \sum_{j \in U_\ell} w_i(j) \cdot (d(\ell, j) + d(j, L))^p \right)^{q/p} \right)^{1/q} \end{aligned}$$

Consider function $\|\cdot\| : \mathbb{R}^m \rightarrow \mathbb{R}$ defined by

$$\|v\| = \left(\sum_{i=1}^n \left(\sum_{j=1}^m w_i(j) \cdot |v_j|^p \right)^{q/p} \right)^{1/q} = \left(\sum_{i=1}^n \left(\sum_{\ell \in K} \sum_{j \in U_\ell} w_i(j) \cdot |v_j|^p \right)^{q/p} \right)^{1/q}.$$

Let Λ_i be a linear map that sends $v \in \mathbb{R}^m$ to $(w_i(1)^{1/p} v_1, \dots, w_i(j)^{1/p} v_j, \dots, w_i(m)^{1/p} v_m)$. Note that $\|v\| = \left\| \|\Lambda_1 v\|_p, \dots, \|\Lambda_n v\|_p \right\|_q$. Therefore, $\|\cdot\|$ is a seminorm on \mathbb{R}^m . In particular,

$$\begin{aligned} \left(\sum_{i=1}^n \left(\sum_{\ell \in K} \sum_{j \in U_\ell} w_i(j) \cdot (d(\ell, j) + d(j, L))^p \right)^{q/p} \right)^{1/q} \\ \leq \left(\sum_{i=1}^n \left(\sum_{\ell \in K} \sum_{j \in U_\ell} w_i(j) \cdot d(\ell, j)^p \right)^{q/p} \right)^{1/q} + \left(\sum_{i=1}^n \left(\sum_{\ell \in K} \sum_{j \in U_\ell} w_i(j) \cdot d(j, L)^p \right)^{q/p} \right)^{1/q} \end{aligned}$$

The second term on the right is simply the cost of L with respect to instance \mathcal{I} and thus equals $\text{cost}_{p,q}^{\mathcal{I}}(L)$. Now, we upper bound the first term. Since every $j \in U_\ell$ is assigned to ℓ , $d(\ell, j) \leq \frac{2}{\gamma^{1/\nu}} C(j)$. Thus the first term is upper bounded by

$$\left(\sum_{i=1}^n \left(\sum_{\ell \in K} \sum_{j \in U_\ell} w_i(j) \cdot \left(\frac{2}{\gamma^{1/\nu}} C(j) \right)^p \right)^{q/p} \right)^{1/q} \stackrel{\text{by (A.2)}}{\leq} \frac{2}{\gamma^{1/\nu}} \left(\sum_{i=1}^n z_i \right)^{1/q} \leq \frac{2}{\gamma^{1/\nu}} B.$$

We conclude that

$$\left(\sum_{i=1}^n \left(\sum_{\ell \in K} w'_i(\ell) \cdot d(\ell, L)^p \right)^{q/p} \right)^{1/q} \leq \text{cost}_{p,q}^{\mathcal{I}}(L) + \frac{2}{\gamma^{1/\nu}} B.$$

Now we construct a proper solution L' for K . For every point in $j \in L$, we choose a closest to j point ℓ in K (breaking ties arbitrarily) and add it to L' . It is immediate that for every $\ell \in K$, $d(\ell, L') \leq 2d(\ell, L)$. Thus,

$$\text{cost}_{p,q}^{\mathcal{I}'}(L') = \left(\sum_{i=1}^n \left(\sum_{\ell \in K} w'_i(\ell) \cdot d(\ell, L')^p \right)^{q/p} \right)^{1/q} \leq 2 \left(\text{cost}_{p,q}^{\mathcal{I}}(L) + \frac{2}{\gamma^{1/\nu}} B \right).$$

As required, we constructed a feasible solution for \mathcal{I}' of cost at most $2\text{cost}_{p,q}^{\mathcal{I}}(L) + \frac{4}{\gamma^{1/\nu}} B$.

Now we prove that inequality (3.22) holds. The proof is almost identical to that of inequality (3.21).

$$\begin{aligned} \text{cost}_{p,q}^{\mathcal{I}}(L) &= \left(\sum_{i=1}^n \left(\sum_{j \in [m]} w_i(j) \cdot d(j, L)^p \right)^{q/p} \right)^{1/q} = \left(\sum_{i=1}^n \left(\sum_{\ell \in K} \sum_{j \in U_\ell} w_i(j) \cdot d(j, L)^p \right)^{q/p} \right)^{1/q} \\ &\leq \left(\sum_{i=1}^n \left(\sum_{\ell \in K} \sum_{j \in U_\ell} w_i(j) \cdot (d(\ell, j) + d(\ell, L))^p \right)^{q/p} \right)^{1/q} \\ &\leq \left(\sum_{i=1}^n \left(\sum_{\ell \in K} \sum_{j \in U_\ell} w_i(j) \cdot d(\ell, j)^p \right)^{q/p} \right)^{1/q} + \left(\sum_{i=1}^n \left(\sum_{\ell \in K} \sum_{j \in U_\ell} w_i(j) \cdot d(\ell, L)^p \right)^{q/p} \right)^{1/q} \end{aligned}$$

As we showed above, the first term is at most $\frac{2}{\gamma^{1/\nu}} B$; the second term equals $\text{cost}_{p,q}^{\mathcal{I}'}(L)$. Therefore,

$$\text{cost}_{p,q}^{\mathcal{I}}(L) \leq \text{cost}_{p,q}^{\mathcal{I}'}(L) + \frac{2}{\gamma^{1/\nu}} B.$$

Now we verify that item 5 holds. Take a point $\ell \in \Lambda$ and point $j \in V_\ell$. Let j' be the closest point to j outside V_ℓ and let ℓ' be the closest point in K to it (so $j' \in V_{\ell'}$). Then we have

$$\begin{aligned} d(j, K \setminus \Lambda) &\leq d(j, \sigma(\ell)) && \text{since } \sigma(\ell) \notin \Lambda \\ &\leq d(j, \ell) + d(\ell, \sigma(\ell)) && \text{by triangle inequality} \\ &\leq d(j, \ell) + d(\ell, \ell') && \text{by choice of } \sigma(\ell) \\ &\leq d(j, \ell) + 2d(j', \ell) && \text{by (A.3)} \\ &\leq d(j, \ell) + 2d(j', j) + 2(j, \ell) && \text{by triangle inequality} \\ &= 3d(j, \ell) + 2d(j, j') = 3d(j, K) + 2d(j, [m] \setminus V_\ell) \end{aligned}$$

Of course, if $\ell \in K \setminus \Lambda$ and $j \in V_\ell$, then $d(j, K \setminus \Lambda) = d(j, K)$. Thus, for *any* center $\ell \in K$ we have

$$d(j, K \setminus \Lambda) \leq 3d(j, K) + 2 \cdot \mathbf{1}_{\ell \in \Lambda} \cdot d(j, [m] \setminus V_\ell).$$

Using that $\|\cdot\|$ is a seminorm as in the proof of item 4, we get

$$\begin{aligned} \text{cost}_{p,q}^{\mathcal{I}}(K \setminus \Lambda) &= \left(\sum_{i=1}^n \left(\sum_{j=1}^m w_i(j) \cdot d(j, K \setminus \Lambda)^p \right)^{q/p} \right)^{1/q} \\ &\leq 3 \left(\sum_{i=1}^n \left(\sum_{j=1}^m w_i(j) \cdot d(j, K)^p \right)^{q/p} \right)^{1/q} + 2 \left(\sum_{i=1}^n \left(\sum_{\ell \in \Lambda} \sum_{j \in V_\ell} w_i(j) \cdot d(j, [m] \setminus V_\ell)^p \right)^{q/p} \right)^{1/q} \\ &= 3\text{cost}_{p,q}^{\mathcal{I}}(K) + 2 \left(\sum_{i=1}^n \left(\sum_{\ell \in \Lambda} \text{vol}_i(V_\ell) \right)^{q/p} \right)^{1/q} \end{aligned}$$

By item 4, the cost of K w.r.t. instance \mathcal{I} , that is, $\text{cost}_{p,q}^{\mathcal{I}}(K)$, is at most the cost of K w.r.t. instance \mathcal{I}' , which is 0, plus $2\gamma^{-1/\nu}B$. We conclude that

$$\text{cost}_{p,q}(K \setminus \Lambda) \leq 6\gamma^{-1/\nu}B + 2\left(\sum_{i=1}^n \left(\sum_{\ell \in \Lambda} \text{vol}_i(V_\ell)\right)^{q/p}\right)^{1/q}.$$

□

Proof. (Proof of Observation 3.1.) Recall that $\sigma(\ell)$ is the closest to ℓ point in K other than ℓ itself. We break ties arbitrarily but consistently so that the set of edges $(\ell, \sigma(\ell))$ forms a forest on K . We choose an arbitrary root in every tree in the forest $(K, (\ell, \sigma(\ell)))$. Then we let K_1 be the set of vertices of odd depth and K_2 be the set of vertices of even depth. Clearly, $\sigma(K_1) \subseteq K_2$ and $\sigma(K_2) \subseteq K_1$. Now,

$$\sum_{\ell \in K} x'_{\ell\sigma(\ell)} = \sum_{\ell \in K} (1 - y_\ell) = |K| - \sum_{\ell \in K} y_\ell = |K| - k.$$

Therefore, $\sum_{\ell \in K_1} x'_{\ell\sigma(\ell)} \geq \frac{|K|-k}{2}$ or $\sum_{\ell \in K_2} x'_{\ell\sigma(\ell)} \geq \frac{|K|-k}{2}$. We are done if the former inequality holds. Otherwise, we simply swap K_1 and K_2 . □

B Connection to Min s -Union and Conjectured Hardness

Let us see how known hardness conjectures for other problems imply polynomial hardness of approximation for (p, q) -Fair Clustering when $q < p - \Omega(1)$.

In the Min s -Union problem, we are given a collection of m sets $S_1, \dots, S_m \subseteq [n]$, and a parameter $s \in [m]$, and wish to find a collection $J \subseteq [m]$ of (indices of) sets of size $|J| = s$ as to minimize the cardinality of their union $|\bigcup_{j \in J} S_j|$. It has been conjectured (see Chlamtáč et al. [2017]) that the following distinguishing problem, slightly rephrased here, is hard (and implies polynomial hardness of approximation for Min s -Union):

CONJECTURE B.1. (DENSE VERSUS RANDOM CONJECTURE) *For any constants $0 < \varepsilon < \delta < 1$, it is hard to distinguish between the following cases when $s \leq \sqrt{m}$:*

- **Random:** $n = m^\varepsilon$ and each of the sets S_1, \dots, S_m is sampled independently by selecting $10 \ln m$ elements in $[n]$ independently at random.
- **Dense:** $n = 10m^\varepsilon \ln m$, and an adversary may select the sets, with an additional guarantee that there exist sets $J \subseteq [m]$ and $I \subseteq [n]$ such that $|J| = s$ and $|I| = (10s \ln m)^\delta$, and for every $j \in J$, $S_j \subseteq I$ (that is, there are s input sets contained in a single set of cardinality $\tilde{O}(s^\delta)$).

This conjecture is supported by Sherali-Adams integrality gaps (cf. Chlamtáč et al. [2017], Chlamtáč and Manurangsi [2018]) and is related to similar conjectures for related problems, and in particular to the Projection Games Conjecture. Bhaskara et al. [2012] show a somewhat weaker (but still polynomial) lower bound for this problem is also supported by a matching integrality gap in the Sum of Squares hierarchy.

A straightforward reduction gives the following:

THEOREM B.1. *For any $1 \leq q \leq p$ and constants $0 < \varepsilon < \delta < 1$, where $\varepsilon \leq \frac{1}{2}$, if the Dense versus Random Conjecture holds for δ, ε , then (p, q) -Fair Clustering is hard to approximate to within less than $m^{\varepsilon(1-\delta)(p-q)/(pq)}$.*

Proof. Let us see a reduction from Min s -Union with these parameters to (p, q) -Fair Clustering, and analyze the optimum clustering value in both cases.

Given an instance (S_1, \dots, S_m, s) of Min s -Union, construct a clustering instance as follows: Identify the points $[m]$ with the sets S_1, \dots, S_m and let the distance between any two points be 1. For all $i \in [n]$ and $j \in [m]$, let $w_i(j) = \mathbb{1}_{i \in S_j}$. Finally, let the target number of centers be $k = m - s$. In the context of the reduction, let us consider instances for which $s = m^\varepsilon (\leq \sqrt{m})$.

Consider first the random case. By a Chernoff bound, w.h.p., for any set $J \subseteq m$ of cardinality $|J| = s = m^\varepsilon$, we have

$$\left| \bigcup_{j \in J} S_j \right| \geq \frac{1}{2} m^\varepsilon \ln n.$$

Thus, for any set of centers K of cardinality $|K| = k$, the cost of K is bounded from below by

$$\begin{aligned} \text{cost}_{p,q}(K) &= \left(\sum_{i=1}^n \left(\sum_{j=1}^m \mathbb{1}_{i \in S_j} d(j, K)^p \right)^{q/p} \right)^{1/q} \\ &= \left(\sum_{i=1}^n |\{j \in [m] \setminus K \mid i \in S_j\}|^{q/p} \right)^{1/q} \\ &\geq \left(\sum_{i=1}^n \mathbb{1}_{i \in \bigcup_{j \in [m] \setminus K} S_j} \right)^{1/q} \\ &= \left| \bigcup_{j \in [m] \setminus K} S_j \right|^{1/q} \geq \left(\frac{1}{2} m^\varepsilon \ln m \right)^{1/q}. \end{aligned}$$

On the other hand, consider the dense case. Let J be the “dense” set as defined in the conjecture, and let $K = [m] \setminus J$. Then as before, the cost of these centers is bounded from above by

$$\begin{aligned} \text{cost}_{p,q}(K) &= \left(\sum_{i=1}^n |\{j \in J \mid i \in S_j\}|^{q/p} \right)^{1/q} \\ &= \left(\sum_{i \in \bigcup_{j \in J} S_j} |\{j \in J \mid i \in S_j\}|^{q/p} \right)^{1/q} \\ &= \left(\sum_{i \in I} |\{j \in J \mid i \in S_j\}|^{q/p} \right)^{1/q} && \text{since } \forall j \in J : S_j \subseteq I \\ &\leq \left(|I|^{(q-p)/p} \left(\sum_{i \in I} |\{j \in J \mid i \in S_j\}| \right)^{q/p} \right)^{1/q} && \text{by Hölder} \\ &\leq |I|^{(q-p)/(pq)} \left(\sum_{j \in J} |S_j| \right)^{1/p} \\ &\leq (10s \ln m)^{\delta(p-q)/(pq)} (10s \ln m)^{1/p} \\ &= (10m^\varepsilon \ln m)^{(\delta(p-q)+q)/(pq)}. \end{aligned}$$

Thus, if the conjecture holds for these parameters, then (p, q) -Fair Clustering cannot be approximated to within less than the ratio between these bounds (the random optimum divided by the dense optimum), which is at least

$$20^{-1/q} (10m^\varepsilon \ln m)^{(1/q) - (\delta(p-q)+q)/(pq)} = 20^{-1/q} (10m^\varepsilon \ln m)^{(1-\delta)(p-q)/(pq)} \geq m^{\varepsilon(1-\delta)(p-q)/(pq)}.$$

□

While not every setting of $0 < \varepsilon < \delta < 1$ with $\varepsilon \leq \frac{1}{2}$ in the conjecture is supported by the same evidence (the Sum of Squares result, for instance, gives a much smaller polynomial gap), it is worth noting that any setting yields a lower bound of $m^{\Omega((p-q)/(pq))}$ for (p, q) -Fair Clustering, and in particular, if ε and δ are arbitrarily close to $\frac{1}{2}$, as is permissible in Conjecture B.1, then we get a lower bound of $m^{(\frac{1}{4}-\eta)\frac{p-q}{pq}}$ for any constant $\eta > 0$.