SIAM J. DISCRETE MATH.
Vol. 26, No. 2, pp. 415–434

ON OPTIMAL WEIGHTED BALANCED CLUSTERINGS: GRAVITY BODIES AND POWER DIAGRAMS*

ANDREAS BRIEDEN† AND PETER GRITZMANN‡

Abstract. We study weighted clustering problems in Minkowski spaces under balancing constraints with a view towards separation properties. First, we introduce the gravity polytopes and more general gravity bodies that encode all feasible clusterings and indicate how they can be utilized to develop efficient approximation algorithms for quite general, hard to compute objective functions. Then we show that their extreme points correspond to strongly feasible power diagrams, certain specific cell complexes, whose defining polyhedra contain the clusters, respectively. Further, we characterize strongly feasible centroidal power diagrams in terms of the local optima of some ellipsoidal function over the gravity polytope. The global optima can also be characterized in terms of the separation properties of the corresponding clusterings.

Key words. clustering, weighted clustering, constrained clustering, power diagram, gravity body, gravity polytope

AMS subject classifications. 91C20, 68H30, 68Q32, 90C27, 68U05

DOI. 10.1137/110832707

1. Introduction. Clustering methods are powerful tools for an abundance of real-world problems; see, e.g., [7]. In the present paper we focus on weighted clustering in Minkowski spaces $(\mathbb{R}^d, \|\,.\,\|)$ under balancing constraints, a problem that is motivated by a new approach to farmland consolidation based on lend-lease agreements; see [11], [12], [9]. Here, a fixed number k of farmers who cultivate a total of m lots want to reduce their operating costs by swapping lots so as to "move" their lots closer together. Of course, each lot may have a different size ω_j. Naturally, the original farm sizes κ_i should not change too much by the reassignment. This means that the new size of the jth farm lies in some interval $[\kappa_i^-, \kappa_i^+]$, where typically $(\kappa_i^+ - \kappa_i^-)/\kappa_i$ is small. Since the focus is not on their geometric shapes, the lots are represented by points in $\mathbb{R}^2$; i.e., in this particular application we have $d = 2$.

Introducing and utilizing the concept of *gravity bodies*, specifically *gravity polytopes*, we will show that each *extremal* (fractional) clustering $\mathcal{C} = (C_1, \dots, C_k)$ admits a *Voronoi dissection* of space; i.e., there exists a polyhedral cell complex whose defining cells $P_1, \dots, P_k$ contain the clusters $C_1, \dots, C_k$, respectively. In fact, the extremal clusterings can be characterized in terms of strongly feasible *power diagrams*; see [1] for a survey on power diagrams. Hence, our results can be seen as a strengthening extension and generalization of those of [6], [4], [18], [19], [8] to the weighted case. Moreover, we identify certain particularly natural power diagrams and study concepts of stability. In particular, we characterize the strongly feasible *centroidal* power diagrams in terms of the local maxima of a convex ellipsoidal function over gravity polytopes. The global maxima correspond to clusterings which maximize a certain measure for the total distance between the clusters.

*Received by the editors May 2, 2011; accepted for publication (in revised form) January 23, 2012; published electronically April 4, 2012.
http://www.siam.org/journals/sidma/26-2/83270.html

†Universität der Bundeswehr, 85579 Neubiberg, Germany (andreas.brieden@unibw.de).
‡Zentrum Mathematik, Technische Universität München, 80290 München, Germany (gritzmann@tum.de).

For background information, closely related results, and further references, see [10], [2], [21], [18], [19], [23], [20], [22], [5], [17], [15] and other papers cited therein.

The present paper is organized as follows. Section 2 gives the basic notation and states our main results. Section 3 defines gravity polytopes and more general gravity bodies as the central objects of the present paper. We give some elementary properties of these bodies and indicate why certain approximations play a crucial role for developing quite efficient approximation algorithms for (weighted) balanced clustering. Section 4 characterizes the extreme points of gravity bodies in terms of strongly feasible power diagrams. Section 5 then links the strongly feasible centroidal power diagrams to the local maxima of an ellipsoidal function over the corresponding gravity polytope. Also the global optima are characterized. Section 6 then closes with some final remarks.

2. Basic notation and main results.

2.1. Constrained clustering. Let, in the following, $d, k, m \in \mathbb{N}$, typically with m much larger than k. To exclude the most trivial cases, we will always assume that $m, k \geq 2$. Let $x_1, \ldots, x_m \in \mathbb{R}^d$ be different points, $\omega_1, \ldots, \omega_m \in]0, \infty[$, $\kappa_1, \ldots, \kappa_k \in]0, \infty[$, with

$$\sum_{i=1}^{k} \kappa_i = \sum_{j=1}^{m} \omega_j.$$

(In the following, intervals will be signified by rectangular brackets; e.g., [,] indicates a closed interval and] , [an open interval.) Further, let $\kappa_1^-, \ldots, \kappa_k^-, \kappa_1^+, \ldots, \kappa_k^+ \in]0, \infty[$ with $\kappa_i^- \leq \kappa_i \leq \kappa_i^+$ for all i, and set $X = \{x_1, \ldots, x_m\}$, $\Omega = (\omega_1, \ldots, \omega_m)$, $\mathrm{K} = (\kappa_1, \ldots, \kappa_k)$, $\mathrm{K}^- = (\kappa_1^-, \ldots, \kappa_k^-)$, $\mathrm{K}^+ = (\kappa_1^+, \ldots, \kappa_k^+)$.

Let $\mathcal{C} = (C_1, \ldots, C_k)$, where $C_i = (\xi_{i,1}, \ldots, \xi_{i,m})$ and $\xi_{i,j} \in [0, 1]$ such that $\sum_{i=1}^{k} \xi_{i,j} = 1$ for each j. Then $\mathcal{C}$ is called a *(fractional) clustering* of X. In fact, $\xi_{i,j}$ is the fraction of x_j assigned to C_i. If all $\xi_{i,j}$ are in $\{0, 1\}$, the clustering $\mathcal{C}$ is called *integer*. If all ω_j are 1, which will be indicated by writing $\Omega = \mathbb{1}$, we will speak of the *combinatorial* case.

The *weight* of the cluster C_i is given by $\omega(C_i) = \sum_{j=1}^{m} \xi_{i,j}\omega_j$; hence each point x_j is counted for C_i with the product of its weight and the fraction belonging to C_i.

A (fractional or integer) clustering is called *balanced* if $\kappa_i^- \leq \omega(C_i) \leq \kappa_i^+$ for each i. These conditions are referred to as *balancing constraints*. If $\mathrm{K}^- = \mathrm{K} = \mathrm{K}^+$, we will speak of *strongly balanced* clusterings; balanced clusterings are then sometimes called *weakly* balanced to emphasize the fact that we allow more general cluster weights.

Now, let $\mathrm{BC}^{\pm}(k, m, X, \Omega, \mathrm{K}^-, \mathrm{K}, \mathrm{K}^+)$ and $\mathrm{BC}(k, m, X, \Omega, \mathrm{K})$ denote the set of all weakly and strongly balanced fractional clusterings for the given parameters, respectively. The set of all such clusterings which are, in addition, integer will be denoted by $\mathrm{BC}_I^{\pm}(k, m, X, \Omega, \mathrm{K}^-, \mathrm{K}, \mathrm{K}^+)$ and $\mathrm{BC}_I(k, m, X, \Omega, \mathrm{K})$, respectively. If the context is clear, we use the abbreviations $\mathrm{BC}^{\pm}$, $\mathrm{BC}_I^{\pm}$, BC, and BC_I. Trivially,

$$\mathrm{BC}_I \subset \mathrm{BC} \subset \mathrm{BC}^{\pm}, \qquad \mathrm{BC}_I \subset \mathrm{BC}_I^{\pm} \subset \mathrm{BC}^{\pm}.$$

Note that if the data Ω, K are integer, the problem of deciding whether $\mathrm{BC}_I^{\pm} \neq \emptyset$ or $\mathrm{BC}_I \neq \emptyset$ is $\mathbb{NP}$-complete [9], while, due to our condition on K, we have $\mathrm{BC} \neq \emptyset$ and $\mathrm{BC}^{\pm} \neq \emptyset$. The latter fact is made explicit by including the parameters $\mathrm{K} = (\kappa_1, \ldots, \kappa_k)$ in the definition of $\mathrm{BC}^{\pm}(k, m, X, \Omega, \mathrm{K}^-, \mathrm{K}, \mathrm{K}^+)$.

One might also wonder why in the notation for this and the other related sets the point set X is mentioned explicitly. Of course, as *combinatorial* objects, the corresponding clusterings do not depend on X. This paper studies, however, *geometric* properties of the "representation" of the clusterings on the set X.

2.2. Separation and dissection. In the following we are interested in separation properties of clusterings. Such properties were studied in [6], [3], [4], [8], and other papers in the strongly balanced (integer) combinatorial case, i.e., for clusterings in $\mathrm{BC}_I(k, m, X, \mathbf{1}, \mathrm{K})$. Some of the results of these papers have been extended to the case of families of sets or, equivalently, to positive integer weights; see, in particular, [18] for such fundamental work which is closely related to our approach.

Let $P_1, \ldots, P_k$ be polyhedra. (Here and throughout the paper it goes without saying that all polyhedra are closed and convex.) $\mathcal{P} = (P_1, \ldots, P_k)$ is a *dissection* of $\mathbb{R}^d$ if $P_1 \cup \cdots \cup P_k = \mathbb{R}^d$ and if the interiors $\mathrm{int}(P_i)$ are pairwise disjoint. $\mathcal{P}$ is a *cell decomposition* of $\mathbb{R}^d$ if $\mathcal{P}$ is a dissection of $\mathbb{R}^d$ and if, for each such choice, the intersection of a face F_i of P_i and a face F_l of P_l is a face of both F_i and F_l. We are interested in cell decompositions of $\mathbb{R}^d$ whose defining polyhedra "contain" the clusters $C_1, \ldots, C_k$ of a given clustering $\mathcal{C}$, respectively. Of course, in the integer case the clusters can be identified with subsets of X, and it is clear what this means. In the general weighted case, we will define this property by using the *support* $\mathrm{supp}(C_i) = \{x_j : \xi_{i,j} \neq 0\}$ of C_i. We say that a cell decomposition $\mathcal{P} = (P_1, \ldots, P_k)$ is *feasible* for $\mathcal{C}$ if $\mathrm{supp}(C_i) \subset P_i$ for all i. In the present paper we will focus on separation properties that are stronger than just feasibility in the following two ways.

Of course, it is clear that in general we have to accept that some points are fractionally assigned to more than one cluster and hence lie in (the boundary of) more than one of the polyhedra of $\mathcal{P}$. (As an example let $k = 2$, $m = 2$, $d = 1$, $x_1 = 0$, $x_2 = 1$, $\omega_1 = 3$, $\omega_2 = 1$, and $\kappa_1 = \kappa_2 = 2$. Then x_1 has to be split among the two clusterings of any strongly balanced clustering.) However, we would like to exclude that points lie in polyhedra "accidentally"; i.e., $x_j \in P_i$ even though $\xi_{i,j} = 0$. Hence, we say that $\mathcal{P}$ *supports* $\mathcal{C}$ if, for all i,

$$\mathrm{supp}(C_i) = X \cap P_i.$$

Note that if $\mathcal{C}$ is integer and $\mathcal{P}$ supports $\mathcal{C}$, then $\mathcal{P}$ is *strictly feasible* for $\mathcal{C}$, i.e., $\mathrm{supp}(C_i) \subset \mathrm{int}(P_i)$ for all i.

The following second condition enforces that $\mathcal{P}$ does not support $\mathcal{C}$ "just by cheating." Before introducing this property more formally, consider the following simple example to indicate what we mean. Let

$$d = k = 2, \quad m = 2, \quad \omega_1 = \omega_2 = 1, \quad \kappa_1 = \kappa_2 = 1, \quad x_1 = -x_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix},$$

and set

$$P_1 = \{(\xi_1, \xi_2)^T \in \mathbb{R}^2 : \xi_1 \leq 0\}, \quad P_2 = \{(\xi_1, \xi_2)^T \in \mathbb{R}^2 : \xi_1 \geq 0\}, \quad \mathcal{P} = (P_1, P_2).$$

Of course, all the balanced clusterings are of the form

$$\mathcal{C}(\delta) = (C_1(\delta), C_2(\delta)) = ((\delta, 1-\delta), (1-\delta, \delta)), \quad (\delta \in [0, 1]).$$

Whenever $\delta \in]0, 1[$, the dissection $\mathcal{P}$ supports $\mathcal{C}(\delta)$. For $\delta \in \{0, 1\}$ this is, however, not the case. On the other hand, the integer clustering $\mathcal{C}(0)$ (or, similarly, $\mathcal{C}(1)$) can be easily transformed into the fractional clustering $\mathcal{C}(\delta)$ for $\delta \in]0,1[$ by simultaneously moving the fraction δ of x_1 from $C_2(0)$ to $C_1(0)$ and the fraction δ of x_2 from $C_1(0)$ to $C_2(0)$. Hence, one can reach the support property of $\mathcal{P}$ by making the clustering "more fractional" and, of course, also worse with respect to many other criteria.

An appropriate general condition can be stated most easily in terms of a labeled multigraph, the *support multigraph* $G(\mathcal{C})$ of the clustering $\mathcal{C} = (C_1, \dots, C_k)$. Its vertex set consists of $C_1, \dots, C_k$, there is an edge between C_i and C_l precisely for every j for which $x_j \in \operatorname{supp}(C_i) \cap \operatorname{supp}(C_l)$, and this edge is labeled with x_j. In the above example, for $\delta \in]0,1[$ the graph $G\big(\mathcal{C}(\delta)\big)$ is a 2-cycle. In fact, the relevant property is essentially that $G(\mathcal{C})$ does not have any cycle. More precisely, a cycle is called *colored* if not all of its labels coincide. (A single-colored clique occurs if a point is split among more than two clusters, e.g., $k = 3$, $m = 1$, $x_1 = 0$, $\omega_1 = 3$, $\kappa_1 = \kappa_2 = \kappa_3 = 1$.) We call $G(\mathcal{C})$ *c-cycle-free* if it does not contain any colored cycle.

While the condition that $G(\mathcal{C})$ be c-cycle-free does not involve $\mathcal{P}$ (and, in fact, expresses that $\mathcal{C}$ corresponds to a vertex of the associated transportation polytope), the following notation is motivated by our interest in clusterings that are induced by dissections of space. Hence, the cell complex $\mathcal{P}$ is called *strongly feasible* for $\mathcal{C}$ if $\mathcal{P}$ supports $\mathcal{C}$ and $G(\mathcal{C})$ is c-cycle-free.

We say that a clustering $\mathcal{C}$ admits a *Voronoi dissection* or has the *Voronoi property* if there is a cell decomposition of $\mathbb{R}^d$ which is feasible for $\mathcal{C}$. If the cell decomposition is strongly (strictly) feasible for $\mathcal{C}$, we refer to it as the *strong (strict) Voronoi property.*

Of particular relevance here are the power diagrams, defined with the aid of different *sites* $s_1, \dots, s_k \in \mathbb{R}^d$ and associated *sizes* $\sigma_1, \dots, \sigma_k \in \mathbb{R}$. (In the standard literature, the σ_i are called weights; we will use the term sizes here to avoid confusion with the weights of the points and clusters we are dealing with.) Specifically, with $S = (s_1, \dots, s_k)$ and $\Sigma = (\sigma_1, \dots, \sigma_k)$ the ith *power cell* $P_i^{S,\Sigma}$ is defined by

$$P_i^{S,\Sigma} = \Big\{ x \in \mathbb{R}^d : \|x - s_i\|_{(2)}^2 - \sigma_i \le \|x - s_j\|_{(2)}^2 - \sigma_j \text{ for all } j \neq i \Big\},$$

where $\|\ .\ \|_{(2)}$ denotes the Euclidean norm. Then $\mathcal{P}^{S,\Sigma} = (P_1^{S,\Sigma}, \dots, P_k^{S,\Sigma})$ is the *power diagram* for (S, Σ). It is easy to see that power diagrams are special cell decompositions of $\mathbb{R}^d$ that generalize the well-known Voronoi diagrams; see [1] for a survey. Hence we can also speak of a power diagram that is (strongly, strictly) feasible for a given clustering.

2.3. Main results. We will characterize the strongly feasible power diagrams in terms of the vertices of certain polytopes that encode our clusterings. To introduce them, again let $\mathcal{C} = \{C_1, \dots, C_k\} \in \mathrm{BC}^{\pm}$, and for $i = 1, \dots, k$ let

$$c_i = c(C_i) = \frac{1}{\omega(C_i)} \sum_{j=1}^{m} \xi_{i,j} \omega_j x_j$$

denote the *center of gravity* of C_i. Note that

$$0 < \kappa_i^- \le \omega(C_i) = \sum_{j=1}^{m} \xi_{i,j} \omega_j \le \kappa_i^+;$$

hence, c_i is well defined. Now,

$$\mathbf{c}(\mathcal{C}) = (c_1^T, \dots, c_k^T)^T$$

is called the *gravity vector* of $\mathcal{C}$. The *gravity body* $Q^{\pm}$ of $\mathrm{BC}^{\pm}$ is then defined as

$$Q^{\pm} = Q^{\pm}(k, m, X, \Omega, \mathrm{K}, \mathrm{K}^{-}, \mathrm{K}^{+}) = \mathrm{conv}\big\{\mathbf{c}(\mathcal{C}) : \mathcal{C} \in \mathrm{BC}^{\pm}\big\}.$$

Other gravity bodies Q, $Q_I^{\pm}$, and Q_I related to BC, $\mathrm{BC}_I^{\pm}$, and BC_I are introduced accordingly. The latter three are polytopes (see Lemma 3.1) and are consequently referred to as *gravity polytopes*. A clustering in any of the defined classes will be called *extremal* if its gravity vector is an extreme point of the corresponding gravity body. Of course, in the polytopal case the extremal clusterings correspond to the vertices of the corresponding gravity polytope.

Now we are ready to state our first main result.

THEOREM 2.1. *$\mathcal{C} \in \mathrm{BC}(k, m, X, \Omega, \mathrm{K})$ is extremal if and only if $\mathcal{C}$ admits a strongly feasible power diagram.*

Note that this theorem can be rephrased by saying that $\mathcal{C}$ is extremal if and only if $G(\mathcal{C})$ is c-cycle-free and there is a power diagram $\mathcal{P}$ that supports $\mathcal{C}$.

As a corollary we see that extreme clusterings lead to strongly feasible power diagrams even in the weakly balanced case.

COROLLARY 2.2. *Let $\mathcal{C} \in \mathrm{BC}^{\pm}(k, m, X, \Omega, \mathrm{K}^{-}, \mathrm{K}, \mathrm{K}^{+})$ be extremal. Then $\mathcal{C}$ admits a strongly feasible power diagram.*

Let us point out that unlike in Theorem 2.1 the converse of Corollary 2.2 does not hold. An example is given in section 4 after the proof of Corollary 4.6.

As we will see, Theorem 2.1 implies, in particular, that in extremal balanced clusterings all but at most $k-1$ points are completely assigned to some cluster.

COROLLARY 2.3. *Let $\mathcal{C} \in \mathrm{BC}^{\pm}(k, m, X, \Omega, \mathrm{K}^{-}, \mathrm{K}, \mathrm{K}^{+})$ be extremal. Then at most $2(k-1)$ variables are fractional.*

Theorem 2.1 will be proved in section 4. For the combinatorial case the characterization follows as a geometric reinterpretation of Theorem 5 of [6]; the integer case with positive integer weights was dealt with in [18]. (In fact, Theorem 5 of [6] and Theorem 3.1 of [18] characterize the vertices of the bounded-shape partition polytopes; see section 3 for a definition.) These papers also include various algorithmic implications (in the binary Turing-machine model). The sufficiency part in the strongly balanced combinatorial case was explicitly given in [4] (see also [3]). More precisely, [4] gave an algorithm that accepts as input (k, m, X, K) and sites $S = (s_1, \dots, s_k)$ and computes a *least-squares clustering* $\mathcal{C} = (C_1, \dots, C_k) \in \mathrm{BC}_I(k, m, X, \mathbb{1}, \mathrm{K})$; i.e., $\mathcal{C}$ minimizes

$$\sum_{i=1}^{k} \sum_{x \in \mathrm{supp}(C_i)} \|s_i - x\|_{(2)}^2$$

among all clusterings in $\mathrm{BC}_I(k, m, X, \mathbb{1}, \mathrm{K})$ and sizes $\Sigma = (\sigma_1, \dots, \sigma_k)$ such that the power diagram $\mathcal{P}^{S,\Sigma}$ is feasible for $\mathcal{C}$. In the plane the running time (in the real RAM model) of this algorithm is computed in [4] to be $O\big(k^2 m \log(m) + km \log^2(m)\big)$ using an optimal space of $O(m)$.

Our second main result characterizes a certain particularly natural class of power diagrams where the sites s_i coincide with the centers c_i of the clusters. More precisely, a power diagram $\mathcal{P}^{S,\Sigma}$ is called *centroidal* for $\mathcal{C}$ if it is feasible for $\mathcal{C}$ and $S = (c_1, \dots, c_k)$. We characterize centroidal power diagrams in terms of the local maxima of the *ellipsoidal function* $\varphi : \mathbb{R}^{kd} \to [0, \infty[$, defined for $z_1, \dots, z_k \in \mathbb{R}^d$ and $\mathbf{z} = (z_1^T, \dots, z_k^T)^T$ by

$$\varphi(\mathbf{z}) = \varphi_{\mathrm{K}}(\mathbf{z}) = \sum_{i=1}^{k} \kappa_i \|z_i\|_{(2)}^2.$$

Note that

$$\| \,.\, \|_{\mathrm{K}} = \big(\varphi(\mathbf{z})\big)^{\frac{1}{2}}$$

is an *ellipsoidal norm*. Hence maximizing φ over a gravity body is a task of *norm maximization*. We say that $\mathcal{C}$ is a *local* (or *global*) *maximizer* of φ in its class if $\mathbf{c}(\mathcal{C})$ is a local (or global) maximizer of φ over the corresponding gravity body.

In order to characterize the strongly feasible centroidal power diagrams, we need one natural condition. Of course, if $\mathrm{supp}(C_i)$ and $\mathrm{supp}(C_l)$ consist of the same single point, then there cannot exist a power diagram that is centroidal for $\mathcal{C}$. Hence we require that the clustering $\mathcal{C}$ be *proper*; i.e., we have for $i \neq l$

$$\big|\mathrm{supp}(C_i)\big| = \big|\mathrm{supp}(C_l)\big| = 1 \quad \Rightarrow \quad \mathrm{supp}(C_i) \neq \mathrm{supp}(C_l).$$

Here is the second main result.

THEOREM 2.4. *Let $\mathcal{C} \in \mathrm{BC}(k, m, X, \Omega, \mathrm{K})$ be proper. Then $\mathcal{C}$ admits a strongly feasible centroidal power diagram if and only if $\mathcal{C} \in \mathrm{BC}(k, m, X, \Omega, \mathrm{K})$ is extremal and a local maximizer for φ.*

As a corollary we see that locally maximal clusterings lead to centroidal power diagrams also in the weakly balanced case.

COROLLARY 2.5. *Let $\mathcal{C} \in \mathrm{BC}^{\pm}(k, m, X, \Omega, \mathrm{K}^-, \mathrm{K}, \mathrm{K}^+)$ be proper and extremal. Further, set $\mathrm{K}(\mathcal{C}) = \big(\omega(C_1), \ldots, \omega(C_k)\big)$, and let $\mathbf{c}(\mathcal{C})$ be a local maximizer for $\varphi_{\mathrm{K}(\mathcal{C})}$. Then $\mathcal{C}$ admits a strongly feasible centroidal power diagram.*

Again, the converse statement does not hold.

Theorem 2.4 will be proved in section 5. There we also discuss the requirement that $\mathcal{C}$ be proper in more detail. Further, we show that the global maximizers of φ correspond to feasible Voronoi dissections that are "most separated" in a certain sense.

3. Gravity bodies.

Before we study some elementary properties of gravity bodies we introduce an approximative variant and indicate why gravity bodies are relevant for practical algorithms. (Naturally, our results have various combinatorial and algorithmic implications that will be dealt with in a broader context in a separate paper.)

Let $\mathcal{C} = \{C_1, \ldots, C_k\} \in \mathrm{BC}^{\pm}$. For $i = 1, \ldots, k$ let

$$\hat{c}_i = \frac{1}{\kappa_i} \sum_{j=1}^{m} \xi_{i,j} \omega_j x_j.$$

Then the point $\hat{c}_i$ can be regarded as an approximation of the center c_i of C_i; it will be referred to as *inexact center* of C_i. Of course, if $\mathcal{C} \in \mathrm{BC}$, then $c_i = \hat{c}_i$.

A natural (and in spite of the $\mathbb{NP}$-hardness of the problem practically quite efficient) approach described in [12] models optimal balanced weighted geometric clustering as a convex maximization problem that involves two norms, a norm $\| \,.\, \|$ on $\mathbb{R}^d$ and some other norm $\| \,.\, \|_\diamond$ on $\mathbb{R}^{k(k-1)/2}$. $\| \,.\, \|_\diamond$ is required to be *monotone*; i.e., $\|x\|_\diamond \leq \|y\|_\diamond$ whenever $x, y \in \mathbb{R}^{k(k-1)/2}$ with $0 \leq x \leq y$. (Here, and in the following, the inequalities are meant componentwise.) Then the convex maximization problem looks as follows:

$$\begin{array}{llcll}
\max & \multicolumn{4}{l}{\left\| \left(\|\hat{c}_1 - \hat{c}_2\|, \|\hat{c}_1 - \hat{c}_3\|, \dots, \|\hat{c}_{k-1} - \hat{c}_k\| \right)^T \right\|_\diamond} \\
\text{subject to} & \kappa_i \hat{c}_i - \sum_{j=1}^m \xi_{i,j}\omega_j x_j & = & 0 & (i = 1,\dots,k), \\
& \sum_{i=1}^k \xi_{i,j} & = & 1 & (j = 1,\dots,m), \\
\kappa_i^- \le & \sum_{j=1}^m \xi_{i,j}\omega_j & \le & \kappa_i^+ & (i = 1,\dots,k), \\
& \xi_{i,j} & \ge & 0 & (i = 1,\dots,k; \\
& & & & j = 1,\dots,m).
\end{array}$$

Here, intuitively, a feasible clustering is optimal if the corresponding inexact centers of gravity are pushed apart as far as possible. Obviously, the convex maximization approach is algorithmically difficult. However, the "hard part" in obtaining optimal clusterings "takes place" only in the $\mathbb{R}^{dk}$ of the k inexact centers. (Note, however, that in the reduced formulation that is obtained by replacing each occurrence of $\hat{c}_i$ by $\frac{1}{\kappa_i}\sum_{j=1}^m \xi_{i,j}\omega_j x_j$ the objective function involves the km variables $\xi_{i,j}$.) One then obtains approximate solutions by approximating the relevant *clustering bodies*

$$C = \left\{ \mathbf{z} = (z_1^T, \dots, z_k^T)^T \in \mathbb{R}^{kd} : \left\| \left(\|z_1 - z_2\|, \|z_1 - z_3\|, \dots, \|z_{k-1} - z_k\| \right)^T \right\|_\diamond \le 1 \right\}$$

in $\mathbb{R}^{dk}$ by polyhedra and solving a linear program in $\mathbb{R}^{km}$ for each of its facets; see [12]. The structure of clustering bodies is studied in [13] in detail, and it is shown that for many choices of norms one can give quite tight worst case bounds for the approximation error. Hence, in spite of the $\mathbb{NP}$-hardness of the general clustering problem one obtains good approximate solutions very efficiently. Note, specifically, that in our prime example of land consolidation $d = 2$ and that k is rather small compared to m. Typically, we have about 8000 variables $\xi_{i,j}$, while k is around 10. This means the hard convex maximization can be approximately dealt with in some $\mathbb{R}^{20}$, while the subsequent (less than 100 different) linear programs take place in some $\mathbb{R}^{8000}$.

Let us further stress this point by introducing inexact variants of our gravity bodies. For $\mathcal{C} = (C_1, \dots, C_k) \in \mathrm{BC}^\pm$ and inexact centers $\hat{c}_1, \dots, \hat{c}_k$, let $\hat{\mathbf{c}}(\mathcal{C}) = (\hat{c}_1^T, \dots, \hat{c}_k^T)^T$ be the *inexact gravity vector* of $\mathcal{C}$, and define the *inexact gravity body* $\hat{Q}^\pm$ of $\mathrm{BC}^\pm$ as

$$\hat{Q}^\pm = \hat{Q}^\pm(k, m, X, \Omega, \mathrm{K}, \mathrm{K}^-, \mathrm{K}^+) = \mathrm{conv}\{\hat{\mathbf{c}}(\mathcal{C}) : \mathcal{C} \in \mathrm{BC}^\pm\}.$$

Other inexact gravity bodies $\hat{Q}$, $\hat{Q}_I^\pm$, and $\hat{Q}_I$ are introduced accordingly.

Note that the above norm maximization task can now be written as

$$\max_{\hat{\mathbf{c}} \in \hat{Q}^\pm} \left\| \left(\|\hat{c}_1 - \hat{c}_2\|, \dots, \|\hat{c}_{k-1} - \hat{c}_k\| \right)^T \right\|_\diamond,$$

which, of course, can be regarded as an approximation of the corresponding maximization over $Q^\pm$. Trivially,

$$Q = \hat{Q}, \qquad Q_I = \hat{Q}_I.$$

Also we have the following simple result.

LEMMA 3.1. *All gravity bodies are compact, and* Q, Q_I, $Q_I^{\pm}$, $\hat{Q}^{\pm}$, $\hat{Q}$, $\hat{Q}_I^{\pm}$, *and* $\hat{Q}_I$ *are polytopes. Further, in the combinatorial case,* $Q^{\pm}$ *is also a polytope, and*

$$Q = Q_I, \qquad Q^{\pm} = Q_I^{\pm}.$$

Proof. Since in the integer case there are only finitely many different clusterings, the bodies Q_I, $Q_I^{\pm}$, $\hat{Q}_I^{\pm}$, and $\hat{Q}_I$ are trivially polytopes.

Let P denote the set of all points $(\xi_{1,1}, \dots, \xi_{1,m}, \dots, \xi_{k,1}, \dots, \xi_{k,m})^T$ that satisfy the constraints

$$0 \le \xi_{i,j}, \quad \sum_{i=1}^{k} \xi_{i,j} = 1, \quad \kappa_i^- \le \sum_{j=1}^{m} \omega_j \xi_{i,j} \le \kappa_i^+$$

for all i and j. Then P is a polytope. Let the function $\psi : P \to \mathbb{R}^{kd}$ map the point $(\xi_{1,1}, \dots, \xi_{1,m}, \dots, \xi_{k,1}, \dots, \xi_{k,m})^T$ to

$$\left(\frac{1}{\sum_{j=1}^{m} \omega_j \xi_{1,j}} \sum_{j=1}^{m} \omega_j \xi_{1,j} x_j, \dots, \frac{1}{\sum_{j=1}^{m} \omega_j \xi_{k,j}} \sum_{j=1}^{m} \omega_j \xi_{k,j} x_j \right).$$

Then ψ is continuous and in the strongly balanced case is, in fact, linear. In the inexact case, the denominators in the definition of ψ are replaced by $\kappa_1, \dots, \kappa_k$, respectively, again leading to a linear function. Hence Q, $\hat{Q}$, and $\hat{Q}^{\pm}$ are polytopes.

To prove the last two equalities, just note that, in the combinatorial case, the conditions defining P are totally unimodular, and hence P is an integer polytope. (See, e.g., [24] for corresponding background information.) □

By Lemma 3.1 we have $Q = Q_I$ and $Q^{\pm} = Q_I^{\pm}$ in the combinatorial case. Hence, in this case, Theorems 2.1 and 2.4 and Corollaries 2.2 and 2.5 can be directly applied to yield strictly feasible (centroidal) power diagrams for extreme integer clusterings. Even if strongly feasible integer clusterings do not exist (as, e.g., in the example with $k = 2$, $m = 2$, $d = 1$, $x_1 = 0$, $x_2 = 1$, $\omega_1 = 3$, $\omega_2 = 1$, and $\kappa_1 = \kappa_2 = 2$ given in subsection 2.2), these results can, nevertheless, be utilized even in the general case to produce *integer* clusterings that admit feasible power diagrams by rounding the (according to Corollary 2.3 typically very few) fractional entries. Of course, the deviation in the balancing constraints increases; however, in the practical problems from farmland consolidation this deviation was small and highly overcompensated by the economic advantages of the new solutions.

Let us point out that in the combinatorial case, the gravity polytopes coincide with the *mean partition polytopes* studied in [14]. Further, as observed in [14], in the strongly balanced combinatorial case, i.e., for $\mathrm{BC}_I(k, m, X, \mathbf{1}, \mathrm{K})$, the gravity polytope Q is just a rescaling of a polytope studied in [6], [16], [19], [20], and other papers there called *single-shape partition polytope*. The more general *bounded-shape partition polytopes* are defined similarly to our gravity bodies with the one difference being that the centers c_i are replaced by the sums

$$\sum_{j=1}^{m} \xi_{i,j} \omega_j x_j;$$

i.e., the division by $\omega(C_i)$ is omitted. The following example shows that in general the set of extreme points of the gravity bodies is richer than that of bounded-shape partition polytopes; i.e., it contains points that may lead to better clusterings.

Let $d = 1$, $k = 2$, $m = 4$, $x_4 = -x_1 = 15$, $x_3 = -x_2 = 3$, $\omega_j = 1$ for all j, and let $\kappa_i = 2$, $\kappa_i^- = 1$, $\kappa_i^+ = 3$. Clearly, the corresponding bounded-shape partition polytope is contained in the linear subspace normal to $(1,1)^T$. In fact, it has the vertices $\pm(-18,18)^T$; the corresponding clusterings are given by $C_1 = (1,1,0,0)$, $C_2 = (0,0,1,1)$ and $C_1' = (0,0,1,1)$, $C_2' = (1,1,0,0)$. Their centers of gravity are $c_1 = \mp 9$, $c_2 = \pm 9$ with Euclidean distance 18. In addition the distance of nearest points of the different clusters, which is an indicator for the quality of separation between the two clusters, is 6. The gravity body $Q_I^{\pm}$, on the other hand, contains the points $\pm(-15,5)$ that are associated with the clusters $C_1 = (1,0,0,0)$, $C_2 = (0,1,1,1)$ and $C_1' = (0,0,0,1)$, $C_2' = (1,1,1,0)$. Here the Euclidean distance of the centers is 20 and the distance of nearest points 12. Hence, we can obtain a clustering that is better with respect to these two measures. However, the corresponding points $\pm(-15,15)$ lie in the relative interior of the bounded-shape partition polytope and will never be maximal with respect to any strictly convex objective function. (As a further advantage in using gravity bodies rather than bounded-shape partition polytopes note that a common translation of the points in X by some vector t results only in a translation of the gravity polytope by the same vector t. Hence all the properties that are relevant here are invariant under common translations of the points of X.)

Let us close this section with the remark that the general objective functions based on $\|\,.\,\|$ and $\|\,.\,\|_\diamond$ include some quite familiar notions. If $\|\,.\,\|$ is the Euclidean norm $\|\,.\,\|_{(2)}$ on $\mathbb{R}^d$ and $\|\,.\,\|_\diamond$ is the ellipsoidal norm on $\mathbb{R}^{k(k-1)/2}$ defined for $v = (\nu_{1,2},\dots,\nu_{k-1,k})^T$ (with coordinates listed in increasing lexicographic order of the index pairs (i,j)) by

$$\|v\|_\diamond = \left(\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\kappa_i\kappa_j\nu_{i,j}^2\right)^{\frac{1}{2}},$$

then

$$\left\|\left(\|c_1-c_2\|,\|c_1-c_3\|,\dots,\|c_{k-1}-c_k\|\right)^T\right\|_\diamond = \left(\sum_{i=1}^{k-1}\sum_{j=i+1}^{k}\kappa_i\kappa_j\|c_i-c_j\|_{(2)}^2\right)^{\frac{1}{2}}.$$

As it turns out, the corresponding maximization problem is equivalent to that involving $\|\,.\,\|_{\mathrm{K}}$. Hence, we are in fact in the case of Theorem 2.4, which concerns the local maxima of $\|\,.\,\|_{\mathrm{K}}$.

Other natural objective functions lead to norm maximization too. For instance, maximizing $\|\,.\,\|_{\mathrm{K}}$ over Q is also equivalent to

$$\min_{\mathcal{C}\in \mathrm{BC}} \quad \sum_{i=1}^{k}\frac{1}{\kappa_i}\sum_{l=1}^{m-1}\sum_{r=l+1}^{m}\xi_{i,l}\omega_l\xi_{i,r}\omega_r\|x_l-x_r\|_{(2)}^2$$

and to

$$\min_{\mathcal{C}\in \mathrm{BC}} \quad \sum_{i=1}^{k}\sum_{j=1}^{m}\xi_{i,j}\omega_j\|c_i-x_j\|_{(2)}^2.$$

Note that, in the combinatorial case, the latter objective function reads as

$$\sum_{i=1}^{k}\sum_{x_j\in \mathrm{supp}(C_i)}\|c_i-x_j\|_{(2)}^2$$

and is, hence, the sum of the “total squared error” within the clusters.

4. Vertices of gravity polytopes and strongly feasible power diagrams. In the following we will prove Theorem 2.1 and Corollary 2.2. One of the key ingredients will be linear programming duality, which has been utilized in the combinatorial case before. We begin with the strongly balanced case.

Let $b_1, \ldots, b_k \in \mathbb{R}^d$ such that $\mathbf{b} = (b_1^T, \ldots, b_k^T)^T$ is in K-*general position*, meaning that

$$\kappa_j b_i \neq \kappa_i b_j \quad (1 \leq i \neq j \leq k).$$

Further, set $a_i = (1/\kappa_i) b_i$ $(i = 1, \ldots, k)$ and

$$a_{i,j} = a_j - a_i = \frac{1}{\kappa_j} b_j - \frac{1}{\kappa_i} b_i \quad (1 \leq i \neq j \leq k).$$

Of course, the vectors $a_{i,j}$ are all nonzero. We study the optimization problem $\max_{\mathbf{c} \in Q} \mathbf{b}^T \mathbf{c}$ in a formulation that does not explicitly involve Q. First note that with

$$\gamma_{i,j} = \omega_j a_i^T x_j$$

for all i, j we have

$$\mathbf{b}^T \mathbf{c} = \sum_{i=1}^{k} b_i^T \left(\frac{1}{\kappa_i} \sum_{j=1}^{m} \omega_j \xi_{i,j} x_j \right) = \sum_{i=1}^{k} \sum_{j=1}^{m} \omega_j \xi_{i,j} a_i^T x_j = \sum_{i=1}^{k} \sum_{j=1}^{m} \xi_{i,j} \gamma_{i,j}.$$

Then we can formulate the optimization problem as the following linear program, which, in fact, is a transportation problem:

$$\begin{array}{lllll}
\text{(LP)} & \max \ \displaystyle\sum_{i=1}^{k} \sum_{j=1}^{m} \gamma_{i,j} \xi_{i,j} & & & \\
\text{subject to} & \displaystyle\sum_{i=1}^{k} \xi_{i,j} & = & 1 & (j = 1, \ldots, m), \\
& \displaystyle\sum_{j=1}^{m} \xi_{i,j} \omega_j & = & \kappa_i & (i = 1, \ldots, k), \\
& \xi_{i,j} & \geq & 0 & (i = 1, \ldots, k; j = 1, \ldots, m).
\end{array}$$

The vector with coefficients $\xi_{i,j}$ (again ordered lexicographically) will be denoted by $(\xi_{i,j})$. Note that, in general, there may be many different clusterings whose gravity vectors coincide. However, the next lemma shows that this is not the case for extremal clusterings. (For a result in the combinatorial case that is closely related to Lemma 4.1 and Theorem 4.7 below, see [6, Corollary 1].)

For $\mathcal{C} \in \mathrm{BC}(k, m, X, \Omega, \mathrm{K})$ let $N(\mathcal{C})$ denote the cone of outer normals of Q in $\mathbf{c} = \mathbf{c}(\mathcal{C})$. Then we have the following result.

LEMMA 4.1. *Let* $\mathcal{C}^* \in \mathrm{BC}(k, m, X, \Omega, \mathrm{K})$ *be extremal, let* $\mathbf{c}^*$ *be its gravity vector, and let* $\mathbf{b} \in \mathrm{int}(N(\mathcal{C}^*))$. *Then* (LP) *has a unique optimum.*

Proof. Let $(\xi^*_{i,j})$ be the optimal solution of (LP) that corresponds to $\mathcal{C}^*$, and suppose that (LP) has a different second optimum $(\xi'_{i,j})$. We consider a subgraph $\mathcal{T}$ of the complete directed bipartite graph on the partition $(\mathcal{C}^*, X)$ of the node set

$\mathcal{C}^* \cup X$. The edge set of $\mathcal{T}$ consists of all edges (C_i^*, x_j) for which $\xi_{i,j}^* > \xi_{i,j}'$ and all edges (x_j, C_i^*) for which $\xi_{i,j}^* < \xi_{i,j}'$; its vertices are those incident to some edge. Each edge (C_i^*, x_j) carries the capacity $\xi_{i,j}^* - \xi_{i,j}'$, while each edge (x_j, C_i^*) has capacity $\xi_{i,j}' - \xi_{i,j}^*$.

Note that at each vertex x_j the sum of the capacities of the ingoing edges is equal to that of the outgoing edges. A similar equation holds for each vertex C_i with the capacities multiplied by the corresponding weights ω_j. Hence $\mathcal{T}$ contains a directed cycle $(C_{i_1}^*, x_{j_1}, \ldots, C_{i_p}^*, x_{j_p}, C_{i_1}^*)$ of some length p.

Now, we consider clusterings $\tilde{\mathcal{C}}$ in $\mathrm{BC}(k, m, X, \Omega, \mathrm{K})$ that are obtained from $\mathcal{C}^*$ by the *cyclic exchange*

$$C_{i_1}^* \xrightarrow{x_{j_1}} C_{i_2}^* \xrightarrow{x_{j_2}} \cdots \xrightarrow{x_{j_{p-1}}} C_{i_p}^* \xrightarrow{x_{j_p}} C_{i_1}^*$$

that simultaneously, for $l = 1, \ldots, p$, with $i_{p+1} = i_1$, moves a fraction δ_{j_l} of the point x_{j_l} from $C_{i_l}^*$ to $C_{i_{l+1}}^*$ such that

$$0 < \delta_{j_l} \leq \xi_{i_l,j_l}^* - \xi_{i_l,j_l}', \qquad \delta_{j_l} \leq \xi_{i_{l+1},j_l}' - \xi_{i_{l+1},j_l}^*, \qquad \omega_{j_l}\delta_{j_l} = \omega_{j_{l+1}}\delta_{j_{l+1}}.$$

Note that, by the third condition, $\alpha = \omega_{j_l}\delta_{j_l}$ is a constant; we call it the *amount* of the cyclic exchange. Of course, by the optimality of $(\xi_{i,j}^*)$ we have

$$\mathbf{b}^T\mathbf{c}^* - \mathbf{b}^T\mathbf{c}(\tilde{\mathcal{C}}) = \sum_{l=1}^{p} (\gamma_{i_{l+1},j_l} - \gamma_{i_l,j_l})\delta_{j_l} \geq 0.$$

On the other hand, since we obtain $(\xi_{i,j}')$ from $(\xi_{i,j}^*)$ by a finite number of such cyclic exchanges, none of these inequalities can be strict; hence

$$\mathbf{b}^T\mathbf{c}^* = \mathbf{b}^T\mathbf{c}(\tilde{\mathcal{C}})$$

for all such $\tilde{\mathcal{C}}$.

So, let $\tilde{\mathcal{C}} = (\tilde{C}_1, \ldots, \tilde{C}_k)$ be some fixed clustering obtained from $\mathcal{C}^*$ by a cyclic exchange, say

$$C_{i_1}^* \xrightarrow{x_{j_1}} C_{i_2}^* \xrightarrow{x_{j_2}} \cdots \xrightarrow{x_{j_{p-1}}} C_{i_p}^* \xrightarrow{x_{j_p}} C_{i_1}^*$$

of amount α, and let $\tilde{\mathbf{c}} = \mathbf{c}(\tilde{\mathcal{C}}) = (\tilde{c}_1^T, \ldots, \tilde{c}_k^T)^T$. Then

$$\tilde{c}_{i_1} = c_{i_1} + \frac{\alpha}{\kappa_{i_1}}(x_{j_p} - x_{j_1}), \qquad \tilde{c}_{i_l} = c_{i_l} + \frac{\alpha}{\kappa_{i_l}}(x_{j_{l-1}} - x_{j_l}) \quad (2 \leq l \leq p),$$

while, of course, the other centers stay the same. Since the points of X are all different, $\tilde{c}_{i_l} \neq c_{i_l}^*$ for all l. Hence $\tilde{\mathbf{c}} \neq \mathbf{c}^*$. On the other hand, $\mathbf{b}^T\tilde{\mathbf{c}} = \mathbf{b}^T\mathbf{c}^*$, contradicting the assumption that c^* is a vertex and $\mathbf{b} \in \mathrm{int}(N(\mathcal{C}^*))$. $\square$

The dual linear program of (LP) is

$$\text{(DLP)} \qquad \min \sum_{i=1}^{k} \kappa_i\mu_i + \sum_{j=1}^{m} \eta_j$$
$$\text{subject to} \qquad \omega_j\mu_i + \eta_j \geq \gamma_{i,j} \qquad (i = 1, \ldots, k; j = 1, \ldots, m).$$

The vector $(\mu_1, \ldots, \mu_k, \eta_1, \ldots, \eta_m)^T$ will be abbreviated as (μ_i, η_j).

Let D denote the feasible region of (DLP). Note that D has 1-dimensional lineality space. However, since the primal program is feasible and has a final optimum, so does the dual. Let F^* denote the optimal face of D, let us indicate a primal-dual pair of optimal solutions by an uppercase $*$ on the variables, and let $\mathcal{C}^* = (C_1^*, \ldots, C_k^*)$ denote the corresponding clustering. Then the complementary slackness conditions read as

$$\xi_{i,j}^*(\omega_j \mu_i^* + \eta_j^* - \gamma_{i,j}) = 0 \qquad (i = 1, \ldots, k; j = 1, \ldots, m).$$

Now, suppose that $\xi_{i,j}^* > 0$ for some index pair i, j. Then, in conjunction with the dual feasibility, it follows that

$$\omega_j \mu_i^* + \eta_j^* - \gamma_{i,j} = 0 \leq \omega_r \mu_l^* + \eta_r^* - \gamma_{l,r}$$

for all index pairs l, r. Specifically, for $j = r$

$$\omega_j(\mu_i^* - \mu_l^*) \leq \gamma_{i,j} - \gamma_{l,j} = \omega_j(a_i^T x_j - a_l^T x_j) = \omega_j a_{l,i}^T x_j.$$

Hence we have

$$\text{supp}(C_i^*) \subset P_i = \bigcap_{l \neq i} \{x : a_{i,l}^T x \leq \mu_l^* - \mu_i^*\}.$$

LEMMA 4.2. *Let* (LP) *have a unique optimum* $(\xi_{i,j}^*)$, *and let the dual optimal point* (μ_i^*, η_j^*) *be contained in the relative interior of the optimal face* F^* *of* D. *Then* $\text{supp}(C_i^*) = X \cap P_i$ *for all* i.

Proof. We show that for all i, j

$$\xi_{i,j}^* \neq 0 \quad \Leftrightarrow \quad \omega_j \mu_i^* + \eta_j^* = \gamma_{i,j}.$$

Let $A = \{(i,j) : \xi_{i,j}^* \neq 0\}$. Then $(\xi_{i,j}^*)_{(i,j) \in A}$ is optimal in the linear program (LP′) that is obtained from (LP) by removing all variables whose index pairs do not belong to A. The dual (DLP′) is then

$$\begin{array}{llccl} (\text{DLP}') & \min & \displaystyle\sum_{i=1}^{k} \kappa_i \mu_i + \sum_{j=1}^{m} \eta_j & & \\ \text{subject to} & \omega_j \mu_i + \eta_j & \geq & \gamma_{i,j} & \big((i,j) \in A\big). \end{array}$$

Let D' denote its feasible region and F' its optimal face. Then F' is given by

$$F' = \bigcap_{(i,j) \in A} \big\{(\mu_i, \eta_j) : \omega_j \mu_i + \eta_j = \gamma_{i,j}\big\}.$$

It suffices to show that F' is the affine hull of F^*. Suppose that this was not the case. Then the cone N of outer normals at D in the point (μ_i^*, η_j^*) would be of higher dimension than the cone N' of outer normals at D', and the dual objective function vector $(\kappa_1, \ldots, \kappa_k, 1, \ldots, 1)^T$ would lie in the relative interior of N. But this implies that there is a solution of (LP) different from $(\xi_{i,j}^*)$. This contradicts the assumption that $(\xi_{i,j}^*)$ is the unique maximizer of (LP). ∎

Note that, in particular, the above proof shows that in the combinatorial case a vertex $(\xi_{i,j}^*)$ of the feasible region of (LP) corresponds to a k-dimensional face of D.

Next we show that the polyhedra $P_1, \dots, P_k$ defined before Lemma 4.2 with the aid of the optimal dual variables μ_i^* actually form a power diagram.

LEMMA 4.3. *Let for all i*

$$s_i = a_i, \quad \sigma_i = \|a_i\|_{(2)}^2 - 2\mu_i^*,$$

and set $S = (s_1, \dots, s_k)$ and $\Sigma = (\sigma_1, \dots, \sigma_k)$. Then for all i

$$P_i = P_i^{S,\Sigma}.$$

Proof. For the proof just note that

$$\|x - s_i\|_{(2)}^2 - \sigma_i \le \|x - s_l\|_{(2)}^2 - \sigma_l$$

is equivalent to

$$2a_{i,j}^T x \le \|a_l\|_{(2)}^2 - \|a_i\|_{(2)}^2 + \sigma_i - \sigma_l = 2(\mu_l^* - \mu_i^*).$$

Hence, $P_i = P_i^{S,\Sigma}$ for all i. ∎

Since it is clear that the corresponding support multigraph is c-cycle-free, Lemmas 4.3, 4.2, and 4.1 imply that a vertex of Q corresponds to a strongly feasible power diagram; i.e., any extreme clustering $\mathcal{C}$ admits a strongly feasible power diagram.

Before we continue with the converse, we prove Corollaries 2.2 and 2.3.

Proof of Corollary 2.2. Let $\mathcal{C}^* \in \mathrm{BC}^{\pm}(k, m, X, \Omega, \mathrm{K}^-, \mathrm{K}, \mathrm{K}^+)$ be extremal, let $\mathbf{c}^* = \big((c_1^*)^T, \dots, (c_k^*)^T\big)^T$ be its center of gravity, and let $\kappa_i^* = \sum_{j=1}^m \omega_j \xi_{i,j}^*$ for all i and $\mathrm{K}^* = (\kappa_1^*, \dots, \kappa_k^*)$. Then $\mathbf{c}^*$ is a vertex of the polytope $Q^* = Q(k, m, X, \Omega, \mathrm{K}^*)$, and the assertion follows from Theorem 2.1. ∎

Proof of Corollary 2.3. The same argument as in the previous proof shows that it suffices to deal with the strongly balanced case.

First note that by Theorem 2.1 the support multigraph $G(\mathcal{C})$ is c-cycle-free. If $G(\mathcal{C})$ is cycle-free, the assertion follows from the fact that a forest on k vertices has at most $k-1$ edges and each edge counts for two fractional variables.

Now suppose there are single-colored cycles. Then, of course, they belong to single-colored maximum cliques in $G(\mathcal{C})$ of size at least 3, called sm-cliques in the following. The sm-cliques come with a "forest-like" structure. More precisely, suppose we would replace each sm-clique by a node and connect two nodes by as many edges as the sm-cliques have vertices in common. Then, since $G(\mathcal{C})$ is c-cycle-free, this graph must be cycle-free.

Now we construct a new graph $\hat{G}(\mathcal{C})$ as follows. We keep all edges that do not belong to an sm-clique but delete in every sm-clique in $G(\mathcal{C})$ all edges except for a spanning tree. Then $\hat{G}(\mathcal{C})$ is cycle-free. Every sm-clique R of size r in $G(\mathcal{C})$ contributes exactly r fractional variables, so each of the $r-1$ edges in the corresponding spanning tree T in $\hat{G}(\mathcal{C})$ contributes $r/(r-1)$, i.e., less than 2, fractional variables. Edges in $\hat{G}(\mathcal{C})$ that do not come from sm-cliques correspond to 2 fractional variables each. Since a forest on k vertices has at most $k-1$ edges, the assertion follows. ∎

Next we prove the converse direction of Theorem 2.1.

LEMMA 4.4. *Let $\mathcal{C} \in \mathrm{BC}(k, m, X, \Omega, \mathrm{K})$, $S := (s_1, \dots, s_k)$, and let $(\xi_{i,j}^*)$ be the feasible solution of* (LP) *associated with $\mathcal{C}$. Further, let $\Sigma = (\sigma_1, \dots, \sigma_k)$, and let $\mathcal{P}^{S,\Sigma}$ be strongly feasible for $\mathcal{C}$. Let for all i*

$$a_i = s_i, \quad \mu_i^* = \frac{1}{2}\big(\|a_i\|_{(2)}^2 - \sigma_i\big),$$

and set

$$\eta_j^* = \gamma_{i,j} - \omega_j \mu_i^*$$

for all i, j *with* $\xi_{i,j}^* \neq 0$*. Then* $(\xi_{i,j}^*)$ *and* (μ_i^*, η_j^*) *are optimal solutions of* (LP) *and* (DLP)*, respectively.*

Proof. First observe that the η_j^* are well defined. In fact, since each point x_j is assigned to some cluster, there is at least one i with $\xi_{i,j}^* \neq 0$. If there is a second such index l, then $x_j \in P_i^{S,\Sigma} \cap P_l^{S,\Sigma}$. This implies

$$\omega_j(\mu_i^* - \mu_l^*) = \omega_j a_{l,i}^T x_j = \omega_j(a_i^T x_j - a_l^T x_j) = \gamma_{i,j} - \gamma_{l,j},$$

and hence

$$\eta_j^* = \gamma_{i,j} - \omega_j \mu_i^* = \gamma_{l,j} - \omega_j \mu_l^*.$$

Next, note that by the definition of η_j^*, the complementary slackness conditions are satisfied. Hence we need only show that (μ_i^*, η_j^*) is feasible for (DLP). Let $x_j \in \mathrm{supp}(C_i)$ and $l \neq i$. Then, by the feasibility of $\mathcal{P}^{S,\Sigma}$ for $\mathcal{C}$, we have

$$0 \leq \|x_j - s_l\|_{(2)}^2 - \sigma_l - \|x_j - s_i\|_{(2)}^2 + \sigma_i = 2a_{l,i}^T x_j + 2\mu_l^* - 2\mu_i^*.$$

Hence

$$\gamma_{l,j} - \gamma_{i,j} = \omega_j a_l^T x_j - \omega_j a_i^T x_j = \omega_j a_{i,l}^T x_j \leq \omega_j \mu_l^* - \omega_j \mu_i^*.$$

Since $\eta_j^* = \gamma_{i,j} - \omega_j \mu_i^*$, we obtain

$$\gamma_{l,j} \leq \omega_j \mu_l^* + \eta_j^*.$$

Thus (μ_i^*, η_j^*) is feasible for (DLP). ∎

The following lemma completes the proof of Theorem 2.1.

LEMMA 4.5. *Let* $\mathcal{C} \in \mathrm{BC}(k, m, X, \Omega, \mathrm{K})$, $S := (s_1, \ldots, s_k)$*, and* $\Sigma = (\sigma_1, \ldots, \sigma_k)$*, and let* $\mathcal{P}^{S,\Sigma}$ *be strongly feasible for* $\mathcal{C}$*. Then* $\mathbf{c} = \mathbf{c}(\mathcal{C})$ *is a vertex of* Q*.*

Proof. Let $(\xi_{i,j}^*)$ be the solution of (LP) associated with $\mathcal{C}$, and let (μ_i^*, η_j^*) be the optimal point of (DLP) defined in Lemma 4.4. Since $G(\mathcal{C})$ is c-cycle-free, there is a unique representation of the objective function vector of (DLP) as a conic combination of the normals of the active constraints of (DLP). Hence $(\xi_{i,j}^*)$ is a vertex of the feasible region of (LP), whence $\mathbf{c}$ is a vertex of Q. ∎

As a corollary, we see that, in the strongly balanced combinatorial case, i.e., for $\mathrm{BC}(k, m, X, \mathbf{1}, \mathrm{K})$, the vertices of the gravity polytope correspond to the strictly feasible power diagrams; see [6], [8] for different proofs of this corollary.

COROLLARY 4.6. *Let* $\mathcal{C} \in \mathrm{BC}(k, m, X, \mathbf{1}, \mathrm{K})$*. Then* $\mathcal{C}$ *is extremal if and only if* $\mathcal{C}$ *admits a strictly feasible power diagram.*

Proof. Let $\mathcal{C}$ be an integer clustering, and let $\mathcal{P}$ be a cell decomposition. Then $\mathcal{P}$ supports $\mathcal{C}$ if and only if $\mathcal{P}$ is strictly feasible for $\mathcal{C}$. The assertion now follows from Theorem 2.1 in conjunction with the fact that, by Lemma 3.1, Q is an integer polytope. ∎

Note that a hyperplane with its normal vector in sufficiently general position supports a polytope in a vertex. Hence, we see that for every feasible power diagram there is another one, whose sites and sizes are arbitrarily close to that of the first, that is, strongly feasible for some clustering of Q.

Naturally, one might wonder whether the full characterization of Theorem 2.1 carries over to the weakly balanced case. However, the following example shows that the converse of Corollary 2.2 does not hold. Let $d = 1$, $k = 2$, $m = 3$, $x_1 = -1$, $x_2 = 0$, $x_3 = 1$, $\omega_1 = \omega_3 = 1$, $\omega_2 = 2$, $\kappa_1 = \kappa_2 = 2$, $\kappa_1^- = \kappa_2^- = 1$, $\kappa_1^+ = \kappa_2^+ = 3$. Then, of course, the clusterings $\mathcal{C}_1 = \big((1,0,0),(0,1,1)\big)$, $\mathcal{C}_2 = \big((0,1,1),(1,0,0)\big)$, $\mathcal{C}_3 = \big((1,1,0),(0,0,1)\big)$, $\mathcal{C}_4 = \big((0,0,1),(1,1,0)\big)$, and $\mathcal{C}_5 = \big((1,1/2,0),(0,1/2,1)\big)$ are feasible. The corresponding gravity vectors are

$$\mathbf{c}_1 = -\mathbf{c}_4 = \frac{1}{3}\begin{pmatrix} -3 \\ 1 \end{pmatrix}, \quad \mathbf{c}_2 = -\mathbf{c}_3 = \frac{1}{3}\begin{pmatrix} 1 \\ -3 \end{pmatrix}, \quad \mathbf{c}_5 = \frac{1}{2}\begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

Note that $\mathbf{c}_5$ is in the interior of $Q^{\pm}$. However, the dissection $\mathcal{P} = (]\infty, 0], [0, \infty[)$, which is, of course, a (centroidal) power diagram, is strongly feasible for $\mathcal{C}_5$.

The above example also shows that a result similar to that of Lemma 4.1 does not hold for bounded-shape partition polytopes. In fact, $(-1,1)^T$ is a vertex of the associated bounded-shape partition polytope but corresponds to $\mathcal{C}_1$, $\mathcal{C}_3$, and $\mathcal{C}_5$. These clusterings are not only different but also behave differently in terms of their separation properties; $\mathcal{C}_1$ and $\mathcal{C}_3$ do have the strict Voronoi property, while $\mathcal{C}_5$ does not. We will close this section by showing that the gravity body behaves differently.

Theorem 4.7. *Let $\mathcal{C}^* = (C_1^*, \ldots, C_k^*)$ and $\mathcal{C}' = (C_1', \ldots, C_k')$ be clusterings of $\mathrm{BC}^{\pm}(k, m, X, \Omega, \mathrm{K}^-, \mathrm{K}, \mathrm{K}^+)$, $\mathbf{c}^* = \mathbf{c}(\mathcal{C}^*)$, $\mathbf{c}' = \mathbf{c}(\mathcal{C}')$, and $\mathbf{b} = (b_1^T, \ldots, b_k^T)^T \in \mathbb{R}^{kd}$. Let $\mathcal{C}^*$ be proper, let $\mathbf{c}^* = \mathbf{c}'$, and suppose that $\mathbf{c}^*$ is the unique maximizer of the linear functional $\mathbf{c} \mapsto \mathbf{b}^T\mathbf{c}$ over $Q^{\pm}$. Then $\mathcal{C}^* = \mathcal{C}'$.*

Proof. The proof is by contraction and, in fact, similar to that of Lemma 4.1. As before, we consider the directed graph $\mathcal{T}$ on the partition $(\mathcal{C}^*, X)$ of the node set $\mathcal{C}^* \cup X$. If $\mathcal{T}$contains a directed cycle, the same argument as in Lemma 4.1 yields a contradiction. Hence we may assume that $\mathcal{T}$ does not contain any such cycle. Of course, this means that $\big(\omega(C_1^*), \ldots, \omega(C_k^*)\big) \neq \big(\omega(C_1'), \ldots, \omega(C_k')\big)$. Now let $(C_{i_1}^*, x_{j_1}, \ldots, C_{i_{p-1}}^*, x_{j_{p-1}}, C_{i_p}^*)$ be a directed path in $\mathcal{T}$. We may assume that $C_{i_1}^*$ has indegree 0 and $C_{i_p}^*$ has outdegree 0. Now, let $\tilde{\mathcal{C}} = (\tilde{C}_1, \ldots, \tilde{C}_k)$ be a clustering in $\mathrm{BC}^{\pm}$ that is obtained from $\mathcal{C}^*$ by the *path exchange*

$$C_{i_1}^* \xrightarrow{x_{j_1}} C_{i_2}^* \xrightarrow{x_{j_2}} \cdots \xrightarrow{x_{j_{p-1}}} C_{i_p}^*$$

of *amount* α that simultaneously, for $l = 1, \ldots, p-1$, moves a fraction δ_{j_l} of the point x_{j_l} from $C_{i_l}^*$ to $C_{i_{l+1}}^*$ such that

$$0 < \delta_{j_l} \leq \xi_{i_l, j_l}^* - \xi_{i_l, j_l}', \qquad \delta_{j_l} \leq \xi_{i_{l+1}, j_l}' - \xi_{i_{l+1}, j_l}^*, \qquad \omega_{j_l}\delta_{j_l} = \omega_{j_{l+1}}\delta_{j_{l+1}}.$$

Let $\tilde{\mathbf{c}} = \mathbf{c}(\tilde{\mathcal{C}}) = (\tilde{c}_1^T, \ldots, \tilde{c}_k^T)^T$. Then, in particular,

$$\tilde{c}_{i_l} = c_{i_l} + \frac{\alpha}{\kappa_{i_l}}(x_{j_{l-1}} - x_{j_l}) \quad (2 \leq l \leq p-1),$$

whence $\tilde{c}_{i_l} \neq c_{i_l}^*$ for all such l. If $p \geq 3$, the same argument as in Lemma 4.1 again yields a contradiction.

So suppose (without loss of generality) that our path exchange is of the form

$$C_1^* \xrightarrow{x_1} C_2^*.$$

If $\tilde{c}_1 \neq c_1^*$ or $\tilde{c}_2 \neq c_2^*$, we again obtain a contradiction. Hence, we have

$$c_1^* = \tilde{c}_1 = \frac{\omega(C_1^*)c_1^* - \alpha x_1}{\omega(C_1^*) - \alpha}, \qquad c_2^* = \tilde{c}_2 = \frac{\omega(C_2^*)c_2^* + \alpha x_1}{\omega(C_2^*) + \alpha},$$

which implies

$$c_1^* = x_1 = c_2^*.$$

By Corollary 2.2, $\mathcal{C}^*$ admits a strongly feasible power diagram. Let H denote the corresponding hyperplane that separates C_1^* from C_2^*. Then, of course, $c_1^*, c_2^* \in H$. This implies $\text{supp}(C_1^*), \text{supp}(C_2^*) \subset H$; thus $\text{supp}(C_1^*) = \text{supp}(C_2^*)$. Since $\mathcal{C}^*$ is proper, this contradicts the fact that the support multigraph $G(\mathcal{C}^*)$ is c-cycle-free. ∎

Note that, in Theorem 4.7, the assumption that $\mathcal{C}^*$ is proper cannot be abandoned.

5. Centroidal power diagrams. In this section, we prove Theorem 2.4. Let us begin with an observation related to the assumption that $\mathcal{C}$ is proper.

LEMMA 5.1. *Let $\mathcal{C} \in \text{BC}$ be proper and extremal, and let $\mathbf{b} = (b_1^T, \ldots, b_k^T)^T \in \text{int}(N(\mathcal{C}))$. Then $\mathbf{b}$ is in* K*-general position, and hence $a_{i,j} \neq 0$ for all i, j.*

Proof. Let $\tilde{\mathcal{C}}$ be obtained from $\mathcal{C}$ by a cyclic exchange

$$C_i \xrightarrow{x_i} C_j \xrightarrow{x_j} C_i$$

of length 2 and some positive amount α, involving two different points x_i and x_j, and let $\tilde{\mathbf{c}}$ denote its center of gravity. Then $\mathbf{c} \neq \tilde{\mathbf{c}}$. Suppose now that $\kappa_j b_i = \kappa_i b_j$. Then

$$\mathbf{b}^T\mathbf{c} - \mathbf{b}^T\tilde{\mathbf{c}} = \frac{\alpha}{\kappa_i} b_i^T(x_j - x_i) + \frac{\alpha}{\kappa_j} b_j^T(x_i - x_j) = 0;$$

hence $\mathbf{b} \notin \text{int}(N(\mathcal{C}))$, a contradiction. ∎

If $\mathcal{C} \in \text{BC}(k, m, X, \Omega, \text{K})$ is extremal and $\mathbf{b} \in \text{int}(N(\mathcal{C}))$, then, by Lemma 5.1, the vector $\mathbf{b}$ is in K-general position if $\mathcal{C}$ is proper. Otherwise there are two clusters C_i, C_j $(i \neq j)$ whose supports consist of the same single point x_0 of $\mathbb{R}^d$, and it might not seem appropriate to distinguish C_i and C_j at all. Rather, one might consider the joint cluster with support $\{x_0\}$ and cluster weight $\kappa_i + \kappa_j$. If one insists, however, one may introduce an arbitrary weakly separating hyperplane through x_0. This leads to a strongly feasible Voronoi dissection which, however, cannot be centroidal.

Now, again let $\text{K} = (\kappa_1, \ldots, \kappa_k)$, set $\text{K}^{-1} = (1/\kappa_1, \ldots, 1/\kappa_k)$, and let

$$\begin{array}{rcll} D & = & \text{diag}(\kappa_1, \ldots, \kappa_1, \ldots, \kappa_k, \ldots, \kappa_k) & \in \mathbb{R}^{(dk)\times(dk)}, \\ D^{\frac{1}{2}} & = & \text{diag}(\sqrt{\kappa_1}, \ldots, \sqrt{\kappa_1}, \ldots, \sqrt{\kappa_k}, \ldots, \sqrt{\kappa_k}) & \in \mathbb{R}^{(dk)\times(dk)}. \end{array}$$

Note that the unit ball $\mathbb{B}_\text{K}$ with respect to $\|\mathbf{z}\|_\text{K}$ and its polar, the unit ball $\mathbb{B}_\text{K}^\circ$ with respect to $\|\mathbf{z}\|_{\text{K}^{-1}}$, can be expressed as

$$\begin{aligned} \mathbb{B}_\text{K} &= \{\mathbf{z} : \|\mathbf{z}\|_\text{K} \leq 1\} = \{\mathbf{z} : \mathbf{z}^T D\mathbf{z} \leq 1\} = \{\mathbf{z} : \varphi(\mathbf{z}) \leq 1\}, \\ \mathbb{B}_\text{K}^\circ &= \mathbb{B}_{\text{K}^{-1}} = \{\mathbf{y} : \mathbf{z} \in \mathbb{B}_\text{K} \Rightarrow \mathbf{y}^T\mathbf{z} \leq 1\} = \{\mathbf{z} : \mathbf{z}^T D^{-1}\mathbf{z} \leq 1\}. \end{aligned}$$

The next lemma is needed in the proof of Theorem 2.4. It is a special case of a more general folklore result on polarity; an elementary proof is included as a service to the reader.

LEMMA 5.2. *Let $\mathbf{z}^* \in \mathbb{R}^{dk}$, $\rho = \|\mathbf{z}^*\|_\text{K}$, and $\mathbf{b}^* = (1/\rho)D\mathbf{z}^*$. Then $\mathbf{b}^*$ spans the cone of outer normals of $\rho\mathbb{B}_\text{K}$ at $\mathbf{z}^*$. Further, $\mathbf{z}^*$ is the unique maximizer of the linear functional $\mathbf{z} \mapsto \mathbf{z}^T\mathbf{b}^*$ over $\rho\mathbb{B}_\text{K}$, and $\mathbf{b}^*$ is the unique maximizer of the linear functional $\mathbf{b} \mapsto \mathbf{b}^T\mathbf{z}^*$ over $\mathbb{B}_\text{K}^\circ$.*

Proof. Let $\mathbf{z} \in \mathbb{B}_{\mathrm{K}} \setminus \{0\}$ and $\mathbf{b} \in \mathbb{B}_{\mathrm{K}}^{\circ} \setminus \{0\}$. Then it follows from the Cauchy–Schwarz inequality that

$$\begin{aligned}\mathbf{b}^T\mathbf{z} = \mathbf{b}^T D^{-\frac{1}{2}} D^{\frac{1}{2}} \mathbf{z} &= (D^{-\frac{1}{2}}\mathbf{b})^T (D^{\frac{1}{2}}\mathbf{z}) \leq \|D^{-\frac{1}{2}}\mathbf{b}\|_{(2)} \cdot \|D^{\frac{1}{2}}\mathbf{z}\|_{(2)} \\ &= \left(\mathbf{b}^T D^{-1}\mathbf{b}\right)^{\frac{1}{2}} \cdot \left(\mathbf{z}^T D\mathbf{z}\right)^{\frac{1}{2}} = \|\mathbf{b}\|_{\mathrm{K}^{-1}} \cdot \|\mathbf{z}\|_{\mathrm{K}} \leq 1\end{aligned}$$

with equality if and only if

$$\|\mathbf{b}\|_{\mathrm{K}^{-1}} = \|\mathbf{z}\|_{\mathrm{K}} = 1 \quad \text{and} \quad \mathbf{b} = D\mathbf{z}.$$

This implies the assertion. ∎

The following lemma proves one direction of Theorem 2.4.

LEMMA 5.3. *Let $\mathcal{C} \in \mathrm{BC}(k, m, X, \Omega, \mathrm{K})$ be proper and extremal, let $\mathbf{c} = c(\mathcal{C}) = (c_1^T, \ldots, c_k^T)^T$ be a local maximizer of $\| \,.\, \|_{\mathrm{K}}$ over Q, and set $S = (c_1, \ldots, c_k)$. Then there exist sizes $\Sigma = (\sigma_1, \ldots, \sigma_k)$ such that the power diagram $\mathcal{P}^{S,\Sigma}$ is strongly feasible for $\mathcal{C}$.*

Proof. First, note that $\mathbf{c} \in \|D^{\frac{1}{2}}\mathbf{c}\|_{(2)}\mathbb{B}_{\mathrm{K}}$. By Lemma 5.2, $D\mathbf{c}$ spans the cone of outer normals at $\|D^{\frac{1}{2}}\mathbf{c}\|_{(2)}\mathbb{B}_{\mathrm{K}}$ in $\mathbf{c}$. Since $\mathbf{c}$ is a local maximizer of $\| \,.\, \|_{\mathrm{K}}$ over Q and extreme, we have $D\mathbf{c} \in \mathrm{int}(N(\mathcal{C}))$. Hence by Lemma 5.1 and Theorem 2.1, there exists $\mathcal{P}^{S,\Sigma}$ that is strongly feasible for $\mathcal{C}$. ∎

The next lemma proves the reverse direction of Theorem 2.4.

LEMMA 5.4. *Let $\mathcal{C} \in \mathrm{BC}(k, m, X, \Omega, \mathrm{K})$, $\mathbf{c} = c(\mathcal{C}) = (c_1^T, \ldots, c_k^T)^T$, and $S = (c_1, \ldots, c_k)$, and let $\mathcal{P}^{S,\Sigma}$ be strongly feasible for $\mathcal{C}$. Then $\mathbf{c}$ is a local maximizer of $\| \,.\, \|_{\mathrm{K}}$ over Q.*

Proof. Since $\mathcal{P}^{S,\Sigma}$ is strongly feasible for $\mathcal{C}$, Lemmas 4.5 and 4.4 imply that $D\mathbf{c} \in \mathrm{int}(N(\mathcal{C}))$. By Lemma 5.2, $D\mathbf{c}$ spans the normal cone of $\|\mathbf{c}\|_{\mathrm{K}}\mathbb{B}_{\mathrm{K}}$ at $\mathbf{c}$. Hence, $\mathbf{c}$ is a local maximizer of $\| \,.\, \|_{\mathrm{K}}$ over Q. ∎

Of course, the proof of Corollary 2.5 now follows with the same argument used in the proof of Corollary 2.2.

Next we characterize the global maxima of φ over Q. We will show that a norm maximal clustering with respect to $\| \,.\, \|_{\mathrm{K}}$ maximizes the *total linear intercluster distance*

$$g(\mathcal{C}, \mathbf{b}) = \sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \sum_{l=1}^{m} \sum_{r=1}^{m} \xi_{i,l}\omega_l \xi_{j,r}\omega_r a_{i,j}^T (x_r - x_l)$$

for $\mathcal{C} \in \mathrm{BC}$ and $\mathbf{b} \in \mathrm{int}(N(\mathcal{C}))$. (To avoid confusion, let us mention that the use of terms may differ in different communities dealing with clustering problems. Here, of course, we refer to a measure to which, in effect, only pairs of points that lie in different clusters can contribute.) In the following, we use the abbreviations

$$\tau = \sum_{j=1}^{m} \omega_j, \qquad \lambda = -\left(\sum_{i=1}^{k} \kappa_i a_i\right)^T \left(\sum_{j=1}^{k} \kappa_j c_j\right).$$

Of course, for strongly balanced clusterings, $\tau = \sum_{i=1}^{k} \kappa_i$ and

$$\sum_{i=1}^{k} \kappa_i c_i = \sum_{j=1}^{m} \omega_j x_j.$$

Note that we could assume $\lambda = 0$ without restricting generality. In fact, any intrinsic property of a clustering is invariant under a common translation of the points $x_1, \dots, x_m$. Hence we may add the translation vector $t = -(1/\tau)\sum_{j=1}^m \omega_j x_j$ to every point of X.

THEOREM 5.5. *Let* $\mathcal{C}, \mathcal{C}^* \in \mathrm{BC}(k, m, X, \Omega, \mathrm{K})$, *let* $\mathbf{c} = (c_1^T, \dots, c_k^T)^T$, *and let* $\mathbf{c}^*$ *be its gravity vectors, respectively, and let* $\mathbf{b} = (b_1^T, \dots, b_k^T)^T \in \mathbb{B}_{\mathrm{K}}^{\circ}$. *Further let* $\mathbf{c}^*$ *be* $\| \,.\, \|_{\mathrm{K}}$*-norm maximal over* Q. *Then*

$$g(\mathcal{C}, \mathbf{b}) = \tau \mathbf{b}^T \mathbf{c} + \lambda \leq \tau \|\mathbf{c}^*\|_{\mathrm{K}} + \lambda$$

with equality if and only if $\mathbf{c}$ *is* $\| \,.\, \|_{\mathrm{K}}$*-norm maximal over* Q *and* $\mathbf{b} = D\mathbf{c}/\|\mathbf{c}^*\|_{\mathrm{K}}$.

Proof. Let $1 \leq i < j \leq k$. Then

$$\begin{aligned}\sum_{l=1}^m \sum_{r=1}^m \xi_{i,l}\omega_l \xi_{j,r}\omega_r a_{i,j}^T (x_r - x_l) &= \kappa_i \sum_{r=1}^m \xi_{j,r}\omega_r a_{i,j}^T x_r - \kappa_j \sum_{l=1}^m \xi_{i,l}\omega_l a_{i,j}^T x_l \\ &= \kappa_i a_{i,j}^T \left(\sum_{r=1}^m \xi_{j,r}\omega_r x_r \right) - \kappa_j a_{i,j}^T \left(\sum_{l=1}^m \xi_{i,l}\omega_l x_l \right) = \kappa_i \kappa_j a_{i,j}^T (c_j - c_i).\end{aligned}$$

Summing up over all pairs of clusters yields

$$\begin{aligned}2g(\mathcal{C}, \mathbf{b}) &= \sum_{i=1}^k \sum_{j=1}^k \kappa_i \kappa_j a_{i,j}^T (c_j - c_i) = \sum_{i=1}^k \sum_{j=1}^k \kappa_i \kappa_j \left(a_j - a_i\right)^T \left(c_j - c_i\right) \\ &= 2\tau \sum_{i=1}^k \kappa_i a_i^T c_i + 2\lambda = 2\lambda + 2\tau \sum_{i=1}^k b_i^T c_i = 2(\tau \mathbf{b}^T \mathbf{c} + \lambda).\end{aligned}$$

The other statements now again follow from the Cauchy–Schwarz inequality. ∎

As a corollary we characterize clusterings with globally maximal total linear inter cluster distance. To state the result precisely, we use again the abbreviation $\mathrm{K}(\mathcal{C}) = \big(\omega(C_1), \dots, \omega(C_k)\big)$ for $\mathcal{C} = (C_1, \dots, C_k)$. For a clearer formulation we will also assume that $\lambda = 0$. (As pointed out before, this is no restriction of generality.)

COROLLARY 5.6. *Let* $\lambda = 0$. *Then the following maxima are attained, and we have*

$$\max_{\mathcal{C} \in \mathrm{BC}^{\pm}} \max_{\mathbf{b} \in \mathbb{B}_{\mathrm{K}(\mathcal{C})}^{\circ}} g(\mathcal{C}, \mathbf{b}) = \tau \max_{\mathcal{C} \in \mathrm{BC}^{\pm}} \|\mathbf{c}(\mathcal{C})\|_{\mathrm{K}(\mathcal{C})}.$$

Proof. Clearly, with $\mathbb{S}_{(2)}$ denoting the Euclidean unit sphere in $\mathbb{R}^{kd}$, we have for every $\mathcal{C} \in \mathrm{BC}^{\pm}$

$$\max_{\mathbf{b} \in \mathbb{B}_{\mathrm{K}(\mathcal{C})}^{\circ}} g(\mathcal{C}, \mathbf{b}) = \max_{\mathbf{b} \in \mathbb{S}_{(2)}} \frac{g(\mathcal{C}, \mathbf{b})}{\|\mathbf{b}\|_{\mathrm{K}(\mathcal{C})^{-1}}}.$$

Since $\mathrm{K}^- > 0$, the functional on the right-hand side is continuous on the compact set

$$\mathrm{BC}^{\pm}(k, m, X, \Omega, \mathrm{K}^-, \mathrm{K}, \mathrm{K}^+) \times \mathbb{S}_{(2)}.$$

Hence, the maximum on the left-hand side of the assertion exists, and we obtain from Theorem 5.5

$$\begin{aligned}&\max\Big\{ g(\mathcal{C}, \mathbf{b}) : \mathcal{C} \in \mathrm{BC}^{\pm}(k, m, X, \Omega, \mathrm{K}^-, \mathrm{K}, \mathrm{K}^+), \mathbf{b} \in \mathbb{B}_{\mathrm{K}(\mathcal{C})}^{\circ} \Big\} \\ &\quad = \max_{\mathrm{K}^- \leq \mathrm{K} \leq \mathrm{K}^+} \max\Big\{ g(\mathcal{C}, \mathbf{b}) : \mathcal{C} \in \mathrm{BC}(k, m, X, \Omega, \mathrm{K}), \mathbf{b} \in \mathbb{B}_{\mathrm{K}(\mathcal{C})}^{\circ} \Big\} \\ &\quad = \tau \max_{\mathrm{K}^- \leq \mathrm{K} \leq \mathrm{K}^+} \max_{\mathbf{c} \in Q(k, m, X, \Omega, \mathrm{K})} \|\mathbf{c}\|_{\mathrm{K}} = \tau \max_{\mathcal{C} \in \mathrm{BC}^{\pm}} \|\mathbf{c}(\mathcal{C})\|_{\mathrm{K}(\mathcal{C})}.\end{aligned}$$

∎

Theorem 5.5 and Corollary 5.6 can be viewed as a characterization of feasible clusterings that are "most separated" with respect to the total linear intercluster distance. Of course, it is simple to define other notions of best separation. For instance, if one asks for maximal stability of a clustering $\mathcal{C}$ with respect to changes of the sites of a corresponding power diagram, one is led to a normal to $\mathbf{c}(\mathcal{C})$ that is "at maximal distance" to the boundary of $N(\mathcal{C})$.

6. Final remarks. In addition to their direct application to clustering, the results of the present paper suggest studying properties of the classical 0-1-partition polytopes which "live" in $\mathbb{R}^{mk}$, in the typically much lower-dimensional space $\mathbb{R}^d$ of the points X themselves. As an example note that for $X = \{1, \ldots, m\}$ and $m = k$, the corresponding 0-1-incidence polytope lies in $\mathbb{R}^{m^2}$ and the gravity polytope (as the natural presentation of the permutahedron) lies in $\mathbb{R}^m$, while $X \subset \mathbb{R}$. Hence we can, in principle, study properties of this 0-1-partition polytope by analyzing dissections of the real line.

Acknowledgments. The authors are indebted to Franz Aurenhammer for valuable comments on a previous version of the present paper. Also we are pleased to acknowledge highly fruitful discussions with Steffen Borgwardt and Uri Rothblum. Further, we would like to thank the anonymous referees for very useful remarks.

REFERENCES

[1] F. Aurenhammer, *Power diagrams: Properties, algorithms and applications*, SIAM J. Comput., 16 (1987), pp. 78–96.

[2] F. Aurenhammer, *Voronoi diagrams—a survey of a fundamental geometric data structure*, ACM Comput. Surveys, 23 (1991), pp. 345–405.

[3] F. Aurenhammer, F. Hoffmann, and B. Aronov, *Minkowski-type theorems and least-squares partitioning*, in Proceedings of the 8th Annual ACM Symposium on Computational Geometry, Berlin, 1992, pp. 350–357.

[4] F. Aurenhammer, F. Hoffmann, and B. Aronov, *Minkowski-type theorems and least-squares clustering*, Algorithmica, 20 (1998), pp. 61–76.

[5] F. Aurenhammer and R. Klein, *Voronoi diagrams*, in Handbook of Computational Geometry, J. Sack and G. Urrutia eds., North–Holland, Amsterdam, 2000, pp. 201–290.

[6] E. R. Barnes, A. J. Hoffman, and U. G. Rothblum, *Optimal partitions having disjoint convex and conic hulls*, Math. Programming, 54 (1992), pp. 69–86.

[7] S. Basu, I. Davidson, and K. Wagstaff, eds., *Constrained Clustering: Advances in Algorithms, Theory, and Applications*, Chapman & Hall/CRC, Boca Raton, FL, 2008.

[8] S. Borgwardt, *A Combinatorial Optimization Approach to Constrained Clustering*, Ph.D. thesis, Technische Universität München, München, Germany, 2010.

[9] S. Borgwardt, A. Brieden, and P. Gritzmann, *Constrained Minimum-k-Star Clustering and its application to the consolidation of farmland*, Oper. Res., 11 (2011), pp. 1–17.

[10] E. Boros and P. L. Hammer, *On clustering problems with connected optima in Euclidean spaces*, Discrete Math., 75 (1989), pp. 81–88.

[11] A. Brieden, *On the Approximability of (Discrete) Convex Maximization and Its Contribution to the Consolidation of Farmland*, Habilitationsschrift, Technische Universität München, München, Germany, 2003.

[12] A. Brieden and P. Gritzmann, *A quadratic optimization model for the consolidation of farmland by means of lend-lease agreements*, in Operations Research Proceedings 2003: Selected Papers of the International Conference on Operations Research (OR 2003), D. Ahr, R. Fahrion, M. Oswald, and G. Reinelt, eds., Springer, Heidelberg, 2004, pp. 324–331.

[13] A. Brieden and P. Gritzmann, *On clustering bodies: Geometry and polyhedral approximation*, Discrete Comput. Geom., 44 (2009), pp. 508–534.

[14] F.-H. Chang, F. K. Hwang, and U. G. Rothblum, *The mean partition problem*, J. Global Optim., 36 (2006), pp. 21–31.

[15] S. Fortune, *Voronoi diagrams and Delaunay triangulations*, in Handbook of Discrete and Computational Geometry, 2nd ed., J. E. Goodman and J. O'Rourke, eds., Chapman & Hall/CRC, Boca Raton, FL, 2004, pp. 513–528.

[16] B. Gao, F. K. Hwang, W. C. W. Li, and U. G. Rothblum, *Partition polytopes over 1-dimensional points*, Math. Program., 85 (1999), pp. 335–362.

[17] F. K. Hwang, J. S. Lee, Y. C. Liu, and U. G. Rothblum, *Sortability of vector partitions*, Discrete Math., 263 (2003), pp. 129–142.

[18] F. K. Hwang, S. Onn, and U. G. Rothblum, *Representations and characterizations of vertices of bounded-shape partition polytopes*, Linear Algebra Appl., 278 (1998), pp. 263–284.

[19] F. K. Hwang, S. Onn, and U. G. Rothblum, *A polynomial time algorithm for shaped partition problems*, SIAM J. Optim., 10 (1999), pp. 70–81.

[20] F. K. Hwang, S. Onn, and U. G. Rothblum, *Linear-shaped partition problems*, Oper. Res. Lett., 26 (2000), pp. 159–163.

[21] M. Inaba, N. Katoh, and H. Imai, *Applications of weighted Voronoi diagrams and randomization to variance-based k-clustering*, in Proceedings of the 10th Annual ACM Symposium on Computational Geometry, 1994, pp. 332–339.

[22] J. Matousek, *On approximate geometric k-clustering*, Discrete Comput. Geom., 24 (2000), pp. 61–84.

[23] R. Ostrovsky and Y. Rabani, *Polynomial time approximation schemes for geometric-k-clustering*, in Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science, pp. 349–358.

[24] A. Schrijver, *Theory of Linear and Integer Programming*, John Wiley and Sons, Chichester, 1986.