# NONNEGATIVE RADIX REPRESENTATIONS FOR THE ORTHANT $R_+^n$

JEFFREY C. LAGARIAS AND YANG WANG

ABSTRACT. Let $A$ be a nonnegative real matrix which is expanding, i.e. with all eigenvalues $|\lambda| > 1$, and suppose that $|\det(A)|$ is an integer. Let $\mathcal{D}$ consist of exactly $|\det(A)|$ nonnegative vectors in $\mathbf{R}^n$. We classify all pairs $(A, \mathcal{D})$ such that every $x$ in the orthant $\mathbf{R}_+^n$ has at least one radix expansion in base $A$ using digits in $\mathcal{D}$. The matrix $A$ must be a diagonal matrix times a permutation matrix. In addition $A$ must be similar to an integer matrix, but need not be an integer matrix. In all cases the digit set $\mathcal{D}$ can be diagonally scaled to lie in $\mathbf{Z}^n$. The proofs generalize a method of Odlyzko, previously used to classify the one–dimensional case.

## 1. INTRODUCTION

For radix expansions to base $b$, the standard digit set $\mathcal{D} = \{0, 1, \ldots, b-1\}$ has the property that every real number $x$ has at least one radix expansion of the form

$$(1.1) \qquad x = \pm \sum_{j=-k}^{\infty} d_j b^{-j}, \ k \in \mathbf{Z} \text{ and all } d_j \in \mathcal{D}.$$

We call digit sets with this property *feasible for base $b$*, following Odlyzko [12]. Knuth [8] raised the question of describing, for base 10, all feasible digit sets $\mathcal{D}$ of size 10. There are indeed feasible digit sets for base 10 other than the standard one, for example, the non–standard digit set $\mathcal{D} = \{0, 1, 2, 3, 4, 50, 51, 52, 53, 54\}$. More generally, one can ask the same question for an arbitrary integer base $b$ with $|b| \geq 2$, see Matula [11].

An important object in studying feasibility of a digit set is the set

$$(1.2) \qquad T(A, \mathcal{D}) = \left\{ \sum_{j=0}^{\infty} A^{-j} \mathbf{d}_j : \text{ all } \mathbf{d}_j \in \mathcal{D} \right\},$$

which is a compact set that satisfies the set–valued functional equation

$$(1.3) \qquad A(T) = \bigcup_{\mathbf{d} \in \mathcal{D}} (T + \mathbf{d}).$$

If $(b, \mathcal{D})$ is feasible then (1.1) gives

$$\mathbf{R} = \pm \left( \bigcup_{j=0}^{\infty} b^j T(b, \mathcal{D}) \right),$$

and this shows that $T(b, \mathcal{D})$ must have positive Lebesgue measure. In terms of $T$ a necessary and sufficient condition for feasibility of a pair $(b, \mathcal{D})$ is that $T$ contains an open interval which has 0 in its closure. This criterion is not easy to check, however.

It is easy to see that if $|\mathcal{D}| < |b|$ then $\mathcal{D}$ cannot be feasible, for (1.3) implies that the Lebesgue measure of $T(b, \mathcal{D})$ is 0. On the other hand, when $|\mathcal{D}| > |b|$ there are many feasible digit sets, and the task of classifying them seems intractable. The most interesting case occurs for feasible digit sets with $|\mathcal{D}| = |b|$, which we call minimal feasible. In this case the representations of real numbers (1.1) using a minimal feasible digit set $\mathcal{D}$ are essentially irredundant. More precisely, for a feasible digit set each real $x$ has only finitely many expansions (1.1), and aside from a set of Lebesgue measure zero, each $x$ has a constant number of representations, this number being 1 or 2, depending on $\mathcal{D}$. If 0 lies in the interior of $T(A, \mathcal{D})$ this number is 2, and it is 1 otherwise. However even the problem of classifying all the minimal feasible digit sets appears difficult, and it currently remains an open problem.

In 1978 Odlyzko gave a complete classification for the special case of nonnegative minimal feasible pairs $(b, \mathcal{D})$. Let $\mathbf{R}_{+} := \{x : \ x \geq 0\}$ and $\mathbf{Z}_{+} := \mathbf{R}^{+} \cap \mathbf{Z}$. Odlyzko [12] proved the following:

**Theorem 1.1** (Odlyzko 1978). *Let $b \in \mathbf{Z}_{+}$ with $b \geq 2$. Suppose that $\mathcal{E}_i$, $0 \leq i \leq k$, are subsets of $\{0, 1, \ldots, b-1\}$ such that*

$$(1.4) \qquad\qquad \mathcal{E}_0 + \mathcal{E}_1 + \cdots + \mathcal{E}_k = \{0, 1, \ldots, b-1\},$$

$$(1.5) \qquad\qquad |\mathcal{E}_0| \cdot |\mathcal{E}_1| \cdots |\mathcal{E}_k| = b,$$

$$(1.6) \qquad\qquad 1 \in \mathcal{E}_0.$$

*Then for any $\alpha > 0$ the digit set*

$$(1.7) \qquad\qquad \mathcal{D} = \alpha \mathcal{E}_0 + \alpha b \mathcal{E}_1 + \cdots + \alpha b^k \mathcal{E}_k$$

*is minimal feasible for base $b$. Conversely, if $\mathcal{D}$ is minimal feasible for base $b$ and consists of $b$ nonnegative elements, then $\mathcal{D}$ is of the form (1.7) for some $\alpha > 0$ and some subsets $\mathcal{E}_0, \mathcal{E}_1, \ldots, \mathcal{E}_k$ of $\{0, 1, \ldots, b-1\}$ satisfying (1.4)–(1.6).*

Odlyzko's proof showed that $T(b, \mathcal{D})$ is then a finite union of intervals of length $\alpha$. In [9] we observed that for $b \geq 2$ a converse result holds, that if $|\mathcal{D}| = b$ and $T(b, \mathcal{D})$ is a finite union of intervals, then for some translate $\mathcal{D}' = \mathcal{D} + x$ the pair $(b, \mathcal{D}')$ is nonnegative feasible.

An important feature of Odlyzko's proof is that it reduces the classification problem to a problem of factoring cyclotomic polynomials into zero–one polynomial factors. All possible zero–one factorizations were determined earlier by Carlitz and Moser [3].

This paper formulates and proves an $n$–dimensional generalization of Theorem 1.1. Let $A \in M_n(\mathbf{R})$ be an expanding matrix and $\mathcal{D} \subset \mathbf{R}^n$. We say the digit set $\mathcal{D}$ is *feasible for base $A$*, or simply $(A, \mathcal{D})$ *is feasible*, if every $\mathbf{x} \in \mathbf{R}^n$ can be represented in the form

$$(1.8) \qquad\qquad \mathbf{x} = Q\Big( \sum_{j=-k}^{\infty} A^{-j} \mathbf{d}_j \Big), \ k \in \mathbf{Z} \text{ and all } \mathbf{d}_j \in \mathcal{D},$$

where $Q$ is of the form $Q = \operatorname{diag}(\pm 1, \pm 1, \ldots, \pm 1)$. We say that $(A, \mathcal{D})$ is *nonnegative* if $A$ is a nonnegative matrix and $\mathcal{D}$ consists of nonnegative vectors.

As in the one–dimensional case, associated to any pair $(A, \mathcal{D})$ is the compact set

$$(1.9) \qquad T(A, \mathcal{D}) = \left\{ \sum_{j=0}^{\infty} A^{-j} \mathbf{d}_j \; : \; \text{all } \mathbf{d}_j \in \mathcal{D} \right\}.$$

It satisfies the set–valued functional equation

$$(1.10) \qquad A(T) = \bigcup_{\mathbf{d} \in \mathcal{D}} (T + \mathbf{d}),$$

and is the attractor of the iterated function system $\{ \varphi_i \; : \; 1 \leq i \leq |\mathcal{D}| \}$ given by

$$\varphi_i(\mathbf{x}) = A^{-1}(\mathbf{x} + \mathbf{d}_i), \; \mathbf{d}_i \in \mathcal{D}.$$

The feasibility condition (1.8) implies that

$$(1.11) \qquad \mathbf{R}^n = \bigcup_{Q = \mathrm{diag}(\pm 1, \ldots, \pm 1)} Q \left( \bigcup_{j=0}^{\infty} A^j \left( T(A, \mathcal{D}) \right) \right),$$

which shows that $T = T(A, \mathcal{D})$ has positive Lebesgue measure. Now (1.10) implies that there are no feasible sets $\mathcal{D}$ with $|\mathcal{D}| < |\det(A)|$. We say that $\mathcal{D}$ is *minimal feasible* if $|\mathcal{D}| = |\det(A)|$; this is the case we consider in this paper. When $(A, \mathcal{D})$ is minimal feasible then $T$ is a self–affine tile in the terminology of Lagarias and Wang [9], [10].

Our object in this paper is to classify those nonnegative $(A, \mathcal{D})$ in $\mathbf{R}^n$ such that $\mathcal{D}$ is a nonnegative minimal feasible digit set. The overall structure of the proofs follow Odlyzko's approach. However some new phenomena appear in dimensions $n \geq 2$, and there are necessarily extra complications in the proofs. A key feature of the proofs is a determination of the structure of the associated tile $T(A, \mathcal{D})$.

More precisely, let $\mathbf{R}^n_+ := \{ [x_1, \ldots, x_n]^T \; : \; \text{all } x_i \geq 0 \}$ denote the nonnegative orthant in $\mathbf{R}^n$ and let $\mathbf{Z}^n_+ := \mathbf{R}^n_+ \cap \mathbf{Z}^n$. It is clear that a nonnegative pair $(A, \mathcal{D})$ is feasible if and only if all $\mathbf{x} \in \mathbf{R}^n_+$ have at least one radix expansion

$$(1.12) \qquad \mathbf{x} = \sum_{j=-k}^{\infty} A^{-j} \mathbf{d}_j, \; k \in \mathbf{Z} \text{ and all } \mathbf{d}_j \in \mathcal{D}.$$

Our first main result asserts that, unlike the one–dimensional case, there are substantial restrictions on the nonnegative matrices $A$ that possess a nonnegative minimal feasible digit set. We show:

**Theorem 1.2.** *Let $A \in M_n(\mathbf{R})$ be a nonnegative and expanding matrix with $|\det(A)| = b$ that has a nonnegative feasible digit set with $|\mathcal{D}| = b$. Then*

$$(1.13) \qquad A = BP,$$

*in which $B$ is a positive diagonal matrix and $P$ is a permutation matrix. Furthermore if $B = \mathrm{diag}(b_1, \ldots, b_n)$ and the permutation $\sigma$ associated to $P$ has cyclic decomposition $\sigma = \tau_1 \tau_2 \cdots \tau_k$ then*

$$(1.14) \qquad b_j^* := \prod_{i \in \tau_j} b_i \in \mathbf{Z} \text{ for } 1 \leq j \leq k$$

*with all $b_j^* \geq 2$. Conversely, for every such $A$ there exists a nonnegative minimal feasible digit set $\mathcal{D}$.*

Theorem 1.2 is derived as a consequence of the stronger Theorem 5.2 proved in Section 5.

The matrix $A$ in (1.13) is always expanding, but $B$ need not always be expanding, see Example 5.1 in Section 5.

In Section 2 we establish the necessary condition $A = BP$ in Theorem 1.2, which implies that $A^k$ must be a diagonal matrix for some $k \geq 1$. To proceed we study the special case of diagonal matrices $B$ and obtain:

**Theorem 1.3.** *Let $B = \mathrm{diag}(b_1, \ldots, b_n)$ be an expanding diagonal matrix, and let $\mathcal{D} \subset \mathbf{R}_+^n$ be a nonnegative minimal feasible digit set for base $B$. Then there exist nonnegative one–dimensional digit sets $\mathcal{D}_1, \ldots, \mathcal{D}_n \subset \mathbf{R}_+^n$, each $\mathcal{D}_i$ is minimal feasible for base $b_i$, such that*

$$(1.15) \qquad \qquad \mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2 \times \cdots \times \mathcal{D}_n.$$

*In particular,*

$$(1.16) \qquad \qquad T(A, \mathcal{D}) = T(b_1, \mathcal{D}_1) \times \cdots \times T(b_n, \mathcal{D}_n).$$

This result combines with Odlyzko's classification of one–dimensional nonnegative digit sets to give a complete classification for nonnegative diagonal matrices. We establish Theorem 1.3 in several steps. First, in Section 3 we show that with a suitable scale change in $\mathcal{D}$ we reduce to the case that $\mathcal{D} \subset \mathbf{Z}^n$, and, more importantly, then show that $T(A, \mathcal{D})$ is a finite union of lattice $n$–cubes (Theorem 3.2). Once this is done, the problem is transformed to questions concerning factorizations of multivariate polynomials with zero–one coefficients into factors of a similar form. For this we prove a multivariate generalization of the criterion of Odlyzko (Theorem 4.1).

Finally, in Section 5 we consider the case of general nonnegative $A$ of the form (1.13). Using the fact that if $(A, \mathcal{D})$ is a nonnegative minimal feasible digit set, we observe in Section 2 that so is $(A^k, \mathcal{D}_{A,k})$ where

$$\mathcal{D}_{A,k} := \left\{ \sum_{i=0}^{k-1} A^j \mathbf{d}_j \: : \: \text{each } \mathbf{d}_j \in \mathcal{D} \right\}.$$

We may choose $A^k$ diagonal and then Theorem 1.3 applies to $(A^k, \mathcal{D}_{A,k})$. We exploit this fact to classify general $(A, \mathcal{D})$ in Theorem 5.2. At the same time we deduce the necessary condition (1.14) and complete the proof of Theorem 1.2 as a corollary.

The assumption of nonnegativity is crucial to all the results of this paper. Without this restriction there are minimal feasible digit set whose associated region $T(A, \mathcal{D})$ has a fractal boundary, or where $T(A, \mathcal{D})$ consists of infinitely many connected components. Some examples can be found in Barnsley [2], Gilbert [4] and Vince [13]. For minimal feasible digit sets the region $T(A, \mathcal{D})$ tiles $\mathbf{R}^n$, such tiles are studied in [1], [4], [5], [7], [9], [10], [13].

We thank A. M. Odlyzko for helpful discussions.

## 2. Nonnegative Feasible Pairs: General Properties

In this section we assume the feasibility of $(A, \mathcal{D})$, so that $|\mathcal{D}| \geq |\det(A)|$. Iterating the functional equation (1.10) yields

$$(2.1) \qquad \qquad A^m(T) = \bigcup_{\mathbf{d} \in \mathcal{D}_{A,m}} (T + \mathbf{d})$$

where

$$(2.2) \qquad \mathcal{D}_{A,m} := \Big\{ \sum_{j=0}^{m-1} A^j \mathbf{d}_j : \text{ each } \mathbf{d}_j \in \mathcal{D} \Big\}.$$

Then for arbitrary digit sets we have:

**Lemma 2.1.** *For any* $m \geq 1$*, the pair* $(A, \mathcal{D})$ *is feasible if and only if the pair* $(A^m, \mathcal{D}_{A,m})$ *is feasible.*

*Proof.* The lemma follows directly from

$$\Big\{ \sum_{j=-k}^{\infty} A^{-j} \mathbf{d}_j : k \in \mathbf{Z}, \text{ all } \mathbf{d}_j \in \mathcal{D} \Big\} = \Big\{ \sum_{j=-k}^{\infty} A^{-mj} \mathbf{d}_j : k \in \mathbf{Z}, \text{ all } \mathbf{d}_j \in \mathcal{D}_{A,m} \Big\}.$$

$\square$

Next, we assume nonnegativity and show:

**Lemma 2.2.** *Suppose that* $(A, \mathcal{D})$ *is nonnegative and feasible. Then* $\mathbf{0} \in \mathcal{D}$*.*

*Proof.* If $\mathbf{0} \notin \mathcal{D}$ then $\mathbf{0}$ cannot have a radix expansion, i.e., for all $k \in \mathbf{Z}$ and $\mathbf{d}_j \in \mathcal{D}$

$$\mathbf{0} \neq \sum_{j=-k}^{\infty} A^{-j} \mathbf{d}_j.$$

This contradicts the feasibility of $(A, \mathcal{D})$. $\square$

The orthant–covering property (1.12) puts a significant restriction on the possible form of $A$, which forms the necessary condition (1.13) in Theorem 1.3.

**Lemma 2.3.** *Suppose that* $(A, \mathcal{D})$ *is nonnegative and feasible. Then* $A = BP$ *where* $B$ *is a nonnegative diagonal matrix and* $P$ *is a permutation matrix. In particular* $A^m$ *is a diagonal matrix for some* $m > 0$*.*

*Proof.* Since $(A, \mathcal{D})$ is nonnegative and feasible,

$$\mathbf{R}_+^n = \Big\{ \sum_{j=-k}^{\infty} A^{-j} \mathbf{d}_j : k \in \mathbf{Z}, \text{ all } \mathbf{d}_j \in \mathcal{D} \Big\}.$$

So $A(\mathbf{R}_+^n) = \mathbf{R}_+^n$ and $A$ must map the $x_i$–axis to some $x_j$–axis for each $1 \leq i \leq n$. Hence $A = BP$ for some diagonal $B$ and permutation matrix $P$.

If $\sigma$ is a permutation, its associated permutation matrix $P_\sigma$ has

$$(P_\sigma)_{i,\sigma(i)} = 1$$

and we let $\sigma$ act on diagonal matrices by

$$\sigma(B)_{i,i} = B_{\sigma(i),\sigma(i)}.$$

Then for any diagonal matrix $B$

$$BP_\sigma = P_\sigma\big(\sigma^{-1}(B)\big).$$

Using this we conclude that $(BP_\sigma)^k = B' P_\sigma^k$ for some diagonal matrix $B'$, which implies that $A^{n!}$ is a diagonal matrix. $\square$

Note that in Lemma 2.3 $A$ must always be an expanding matrix for feasible digit sets to exist, but, the matrix $B$ in Lemma 2.3 is not necessarily expanding, see Example 5.1 in Section 5.

Lemma 2.1 and Lemma 2.3 reduce the study of nonnegative feasible pairs $(A, \mathcal{D})$ to those for diagonal matrices, except that we must resolve which digit sets on $A^m$ are of the form $\mathcal{D}_{A,m}$ for some digit set $\mathcal{D}$ on $A$. We accomplish this in Section 5 for nonnegative minimal feasible digit sets, after first classifying the allowable form of $\mathcal{D}_{A,m}$ when $A^m$ is diagonal in Section 4.

Next, we suppose that $(A, \mathcal{D})$ is a nonnegative minimal feasible pair, and we deduce some facts concerning $T = T(A, \mathcal{D})$. The relation (1.12) implies

$$(2.3) \qquad \mathbf{R}_+^n = \bigcup_{m=0}^{\infty} A^m \big( T(A, \mathcal{D}) \big),$$

so $T$ has positive Lebesgue measure $\mu(T) > 0$. Now $\mathcal{D}_{A,m}$ has cardinality

$$(2.4) \qquad |\mathcal{D}_{A,m}| \le |\mathcal{D}|^m = |\det(A)|^m,$$

and taking the Lebesgue measure of both sides of (2.1), using $\mu(T) > 0$ implies that $|\mathcal{D}_{A,m}| = |\det(A)|^m$ and the *measure–disjointness property*

$$(2.5) \qquad \mu\big((T + \mathbf{d}_1) \cap (T + \mathbf{d}_2)\big) = 0 \text{ if } \mathbf{d}_1, \mathbf{d}_2 \in \mathcal{D}_{A,m}, \quad \mathbf{d}_1 \ne \mathbf{d}_2.$$

Lemma 2.2 now shows that $\mathbf{0} \in \mathcal{D}$, so we have

$$(2.6) \qquad \mathcal{D} = \mathcal{D}_{A,1} \subseteq \mathcal{D}_{A,2} \subseteq \mathcal{D}_{A,3} \subseteq \cdots.$$

Now set

$$(2.7) \qquad \mathcal{D}_{A,\infty} = \bigcup_{m=1}^{\infty} \mathcal{D}_{A,m}.$$

In view of (2.6) the measure–disjointness property extends to

$$(2.8) \qquad \mu\big((T + \mathbf{d}_1) \cap (T + \mathbf{d}_2)\big) = 0 \text{ for } \mathbf{d}_1, \mathbf{d}_2 \in \mathcal{D}_{A,\infty}, \quad \mathbf{d}_1 \ne \mathbf{d}_2.$$

Furthermore the relation (2.3) now becomes

$$(2.9) \qquad \mathbf{R}_+^n = \bigcup_{\mathbf{d} \in \mathcal{D}_{A,\infty}} (T + \mathbf{d}).$$

By measure–disjointness this says that the orthant $\mathbf{R}_+^n$ is perfectly tiled with copies of $T$, centered at points of $\mathcal{D}_{A,\infty}$. (This is actually a self–replicating tiling of $\mathbf{R}^n$ in the sense of Kenyon [6], [7].) For this reason we call $T(A, \mathcal{D})$ the *tile associated to $\mathcal{D}$*.

## 3. Diagonal Case: Structure of Tile

Suppose that $B$ is a nonnegative diagonal matrix that is expanding. There is then considerable freedom to rescale the digit set $\mathcal{D}$. For any positive real factors $s_1, s_2, \ldots, s_n$ set $S = \text{diag}(s_1, s_2, \ldots, s_n)$ and define

(3.1)
$$S(\mathcal{D}) := \Big\{ \mathbf{d}' = S\mathbf{d} = [s_1 d_1, s_2 d_2, \ldots, s_n d_n]^T : \ \mathbf{d} = [d_1, d_2, \ldots, d_n]^T \in \mathcal{D} \Big\},$$

If $(B, \mathcal{D})$ is feasible, then so is $(B, S(\mathcal{D}))$ and vice versa, since

$$(3.2) \qquad T\big(B, S(\mathcal{D})\big) = S\big(T(B, \mathcal{D})\big).$$

This equality is a consequence of (1.9) because $S$ commutes with $B$.

**Lemma 3.1.** *Let $B = \mathrm{diag}(b_1, b_2, \ldots, b_n)$ and suppose that $(B, \mathcal{D})$ is nonnegative and minimal feasible. For any subset $I \subseteq \{1, 2, \ldots, n\}$ let $\mathcal{D}_I$ denote the subset of the digit set $\mathcal{D}$ consisting of all digits $\mathbf{d}$ whose $j$–th coordinate $d_j = 0$ for all $j \notin I$. Then*

$$(3.3) \qquad |\mathcal{D}_I| = \prod_{i \in I} b_i .$$

*In particular all $b_i$ are integers. If $B_I := \mathrm{diag}(b_i : i \in I)$ and $\mathcal{D}_I$ is re–interpreted as a set of vectors in $\mathbf{R}^{|I|}$ by dropping coordinates outside $I$, then $(B_I, \mathcal{D}_I)$ is minimal feasible in $\mathbf{R}^{|I|}$.*

*Proof.* We use the radix expansion (1.12) confined to the $I$–face of $\mathbf{R}_+^n$, which is

$$(3.4) \qquad \mathbf{R}_+^I := \left\{ [x_1, x_2, \ldots, x_n]^T \; : \; \text{all } x_j \geq 0 \text{ and } x_j = 0 \text{ if } j \notin I \right\}.$$

Since $B$ is diagonal, all $\mathbf{x} \in \mathbf{R}_+^I$ can be represented by a radix expansion (1.12) using digits in $\mathcal{D}_I$. Thus $\bigcup_{k \geq 0} B_I^k \left( T(B_I, \mathcal{D}_I) \right)$ covers $\mathbf{R}_+^{|I|}$, hence $(B_I, \mathcal{D}_I)$ is feasible in $\mathbf{R}^I$. Also $T(B_I, \mathcal{D}_I)$ has positive $|I|$–dimensional Lebesgue measure, which forces

$$|\mathcal{D}_I| \geq \det(B_I) = \prod_{i \in I} b_i .$$

To show that equality occurs, we consider all $|\mathcal{D}_I|^k$ representatives

$$\mathcal{D}_{B,k}^I = \left\{ \sum_{j=0}^{k-1} B^j \mathbf{d}_j \; : \; \text{all } \mathbf{d}_j \in \mathcal{D}_I \right\}.$$

Then for any $\mathbf{f} = [f_1, \ldots, f_n]^T \in \mathcal{D}_{B,k}^I$, its $i$–th coordinate $f_i$ for $i \in I$ satisfies

$$|f_i| = \left| \sum_{j=0}^{k-1} b_i^j d_{j,i} \right| < C b_i^k$$

in which $C = \max_{\mathbf{d} \in \mathcal{D}} |\mathbf{d}|_\infty$ is a constant. So there can be at most $\prod_{i \in I} (C b_i^k) = C^{|I|} (\prod_{i \in I} b_i)^k$ different elements in $\mathcal{D}_{B,k}^I$. By choosing $k$ sufficiently large we obtain $|\mathcal{D}_I| \leq \prod_{i \in I} b_i$. This proves (3.4). The fact that $b_i \in \mathbf{Z}$ follows by choosing $I = \{i\}$ and, the feasibility of $(B_I, \mathcal{D}_I)$ follows from the feasibility of $(B, \mathcal{D})$. $\qquad\square$

**Theorem 3.2.** *Suppose that $B = \mathrm{diag}(b_1, \ldots, b_n)$ with all $b_i > 1$, and that $\mathcal{D}$ is a nonnegative digit set. Then $(B, \mathcal{D})$ is minimal feasible if and only if $|\mathcal{D}| = |\det(B)|$ and there exist positive scaling factors $s_1, s_2, \ldots, s_n$ with $S = \mathrm{diag}(s_1, s_2, \ldots, s_n)$ such that the scaled digit set $\mathcal{D}^* = S(\mathcal{D})$ has the following two properties:*

(i) *$\mathcal{D}^* \subset \mathbf{Z}^n$, and furthermore $\mathcal{D}^*$ contains all $2^n$ zero–one vectors. (A zero–one vector is a vector whose entries are all 0 or 1.)*

(ii) *$T(B, \mathcal{D}^*)$ is a finite union of lattice cubes, i.e. there exists a finite set $\mathcal{E} \subset \mathbf{Z}_+^n$ containing $\mathbf{0}$ such that*

$$(3.5) \qquad T(B, \mathcal{D}^*) = \bigcup_{\mathbf{f} \in \mathcal{E}} (\mathbf{f} + [0, 1]^n).$$

*Moreover, $\left( \mathcal{E} - \mathcal{E} \right) \cap \left( \mathcal{D}_{B,\infty}^* - \mathcal{D}_{B,\infty}^* \right) = \{\mathbf{0}\}$.*

*For any such $(B, \mathcal{D})$ the scaling factors $S$ for which (i), (ii) hold are unique.*

*Proof.* First we show that properties (i), (ii) are sufficient. According to (ii), $[0, 1]^n \subseteq T(B, \mathcal{D}^*)$. Now (3.2) gives

$$[0, s_1^{-1}] \times \cdots \times [0, s_n^{-1}] \subseteq T(B, \mathcal{D}),$$

whence

$$\mathbf{R}_+^n \subseteq \bigcup_{m=1}^{\infty} B^m \big([0, s_1^{-1}] \times \cdots \times [0, s_n^{-1}]\big) \subseteq \bigcup_{m=1}^{\infty} B^m \big(T(B, \mathcal{D})\big).$$

This shows that $\mathcal{D}$ is feasible, and it is minimal by hypothesis.

Conversely, suppose $(B, \mathcal{D})$ is minimal feasible. We prove the existence and uniqueness of $S$ such that (i), (ii) hold by induction on the dimension $n$. The base case $n = 1$ was established by Odlyzko ([12], Lemma 5, and his equation (3.3)). The assertion

$$\big(\mathcal{E} - \mathcal{E}\big) \cap \big(\mathcal{D}_{B, \infty}^* - \mathcal{D}_{B, \infty}^*\big) = \{0\}$$

is a consequence of the measure–disjointness property (2.8).

Suppose the theorem is true for dimensions up to $n - 1$. We consider the sets $I_i = \{1, 2, \ldots, n\} \setminus \{i\}$ and apply the induction hypothesis on each of the $(n-1)$–dimensional orthants $\mathbf{R}_+^{I_i}$ bounding $\mathbf{R}_+^n$, to $(B_{I_i}, \mathcal{D}_{I_i})$, which we may do by Lemma 3.1. In each case we get unique scaling factors $\big(s_1^{(i)}, \ldots, \hat{s}_i^{(i)}, \ldots, s_n^{(i)}\big)$ where $\hat{s}_i^{(i)}$ is omitted, and these rescale the attractors $T(B_{I_i}, \mathcal{D}_{I_i})$ so that each is a finite union of disjoint $(n-1)$–dimensional unit cubes. Furthermore they must be consistent with each other where they are both defined, i.e., $s_j^{(i)} = s_j^{(k)}$ whenever $i, k \neq j$. So we have a unique set of scaling factors $(s_1, s_2, \ldots, s_n)$, $s_i = s_i^{(k)}$ for any $k \neq i$, such that $\mathcal{D}^* = S(\mathcal{D})$ contains $\mathbf{0}$ and all zero–one vectors except possibly

$$\mathbf{e}_1 + \mathbf{e}_2 + \cdots + \mathbf{e}_n = [1, 1, \ldots, 1]^T.$$

Also, the induction hypothesis implies that any digit $\mathbf{d} \in \mathcal{D}^*$ containing a zero entry necessarily lies in $\mathbf{Z}^n$. We do not yet know that $\mathcal{D}^* \subset \mathbf{Z}^n$, however.

We proceed by a series of claims.

*Claim* 1. $T = T(B, \mathcal{D}^*)$ contains the unit cube $[0, 1]^n$.

Suppose not. If all nonzero digits $\mathbf{d} \in \mathcal{D}^*$ have some coordinate at least 1, then all $|B^k \mathbf{d}|_\infty \geq 1$ since $B$ is diagonal. Thus the only expansion (1.12) having $\mathbf{x} \in (0, 1)^n$ would have $k < 0$, so $[0, 1]^n \subseteq T$. Hence there must be some digit $\mathbf{z} = [z_1, \ldots, z_n]^T \in \mathcal{D}^*$ with $0 < z_i < 1$ for all $i$. (If some $z_i = 0$ then we know $\mathbf{z} \in \mathbf{Z}^n$.) For each $1 \leq i \leq n$ we denote

$$w_i = \min_{\mathbf{f} \in \mathcal{D}^* \setminus \mathbf{Z}^n} \Big\{ f_i : \ f = [f_1, \ldots, f_n]^T \Big\}.$$

So $0 < w_i < 1$ for all $i$. This implies that for any $\mathbf{0} \neq \mathbf{d} \in \mathcal{D}^*$, all $B^k \mathbf{d}$ with $k \geq 0$ either give vectors outside the unit cube or else have the $i$–th coordinate at least $w_i$, hence $T$ must contain the slab

$$S_i = [0, 1]^{i-1} \times [0, w_i] \times [0, 1]^{n-i},$$

so

$$\bigcup_{i=1}^{n} S_i \subseteq T.$$

Now $\mathbf{e}_2 \in \mathcal{D}^*$ and we compare $T+\mathbf{e}_2$ with $T+\mathbf{z}$ where $\mathbf{z} \in \mathcal{D}^*$ with all its coordinates $0 < z_i < 1$. We get a contradiction by showing that

$$(3.6) \qquad \mu\Big((T + \mathbf{z}) \cap (T + \mathbf{e}_2)\Big) > 0,$$

contradicting the measure–disjointness property (2.8). To show (3.6) we need only to observe that

$$T + \mathbf{e_2} \supseteq S_1 + \mathbf{e_2} = [0, 1] \times [1, 1 + w_2] \times [0, 1]^{n-2},$$

$$T + \mathbf{z} \supseteq S_2 + \mathbf{z} = [z_1, w_1 + z_1] \times [z_2, z_2 + 1] \times \cdots \times [z_n, z_n + 1].$$

Since $0 < z_i$, $w_i < 1$ for all $i$, both $T + \mathbf{z}$ and $T + \mathbf{e}_2$ contain the small cube $[z_1, 1, \ldots, 1]^T + \varepsilon[0, 1]^n$ for some sufficiently small $\varepsilon$, which establishes (3.6). Thus Claim 1 follows.

*Claim 2.* $\mathbf{e} := \mathbf{e}_1 + \mathbf{e}_2 + \cdots + \mathbf{e}_n = [1, 1, \ldots, 1]^T \in \mathcal{D}^*$.

Suppose not. Now, $[0, 1]^n \subseteq T$, so $B(T) \supseteq [0, 2]^n$. Using the self–affine property

$$(3.7) \qquad B(T) = \bigcup_{\mathbf{d} \in \mathcal{D}^*} (T + \mathbf{d})$$

and the property that $T + \mathbf{d}$ and $T$ are measure–disjoint for all $\mathbf{d} \in \mathcal{D}^*$, there must be a small cube

$$(3.8) \qquad \mathbf{e} + \varepsilon[0, 1]^n \subseteq T.$$

To see this, note that at least one $T+\mathbf{d}$ covers $\mathbf{e}+\varepsilon[0, 1]^n$ when $\varepsilon$ is sufficiently small. For this to happen, $\mathbf{d} = [d_1, \ldots, d_n]^n$ must satisfy $d_i \leq 1 + \varepsilon$ for all $i$. Now none of the translated tiles $T + \mathbf{d}$, $\mathbf{0} \neq \mathbf{d} \in \mathcal{D}^*$ contains any interior point of $\mathbf{e} + \varepsilon[0, 1]^n$, for if they did then $T$ would overlap some $\mathbf{d} + T$, contradicting measure–disjointness. Hence (3.8) holds.

Now we obtain a contradiction using the self–affine property (3.7). The tile $T$ contains $[0, 1]^n$, so $\partial T$ must contain all $(n - 1)$–dimensional faces of the unit cube $[0, 1]^n$, for if not $T$ would overlap at least one translated tile $T+\mathbf{d}$ for some zero–one digit $\mathbf{0} \neq \mathbf{d} \in \mathcal{D}^*$. So the boundary of $B(T)$ contains all $(n - 1)$–dimensional faces of $[0, b_1] \times \cdots \times [0, b_n]$. But the tile $T$ has the unit cube with at least a small cube $\mathbf{e} + \varepsilon[0, 1]^n$ attached to it, and $\bigcup_{\mathbf{d} \in \mathcal{D}^*}(T + \mathbf{d})$ can never produce the "upper" flat face $[0, b_1] \times \cdots \times [0, b_{n-1}] \times \{b_n\}$, because the projecting piece ruins it. Thus $B(T)$ cannot be tiled by translates of $T$, a contradiction that proves Claim 2.

*Claim 3.* $T$ is a finite union of lattice cubes and $\mathcal{D}^* \subset \mathbf{Z}^n$.

We assign a total ordering to $\mathbf{Z}_+^n = \mathbf{Z}^n \cap \mathbf{R}_+^n$ with the property that if $|\mathbf{g}|_1 < |\mathbf{g}'|_1$ then $\mathbf{g} \prec \mathbf{g}'$ in the ordering, where $|\mathbf{x}|_1 := \sum_{i=1}^n |x_i|$ for any $\mathbf{x} \in \mathbf{R}^n$. (There are many such orderings, and all we need is one of them.) We prove the following hypothesis by induction on $\mathbf{g} \in \mathbf{Z}_+^n$:

(i) $\mu\big(T \cap (\mathbf{g} + [0, 1)^n)\big) > 0$ implies $\mathbf{g} + [0, 1)^n \subseteq T$.

(ii) $\mathcal{D}^* \cap (\mathbf{g} + [0, 1)^n) \subset \mathbf{Z}^n$.

For $\mathbf{g} = \mathbf{0}$, the smallest element in $\mathbf{Z}_+^n$ with respect to the ordering, the hypothesis is clearly true: we have $[0, 1)^n \subseteq T$ and $[0, 1)^n \cap \mathcal{D}^* = \{\mathbf{0}\}$, for if $\mathbf{0} \neq \mathbf{d} \in [0, 1)^n \cap \mathcal{D}^*$ we would have $\mu\big(T \cap (\mathbf{d} + T)\big) > 0$ which contradicts measure–disjointness.

Suppose that the induction hypothesis is true for all $\mathbf{g}' \prec \mathbf{g}$. To prove it for $\mathbf{g}$ we argue by contradiction. So suppose that the hypothesis is false for $\mathbf{g}$, so either $\mu\big(T \cap (\mathbf{g} + [0, 1)^n)\big) > 0$ but $\mathbf{g} + [0, 1)^n \not\subseteq T$, or $\mathcal{D}^* \cap (\mathbf{g} + [0, 1)^n) \not\subset \mathbf{Z}^n$.

If $T \supseteq \mathbf{g} + [0,1)^n$ but there exists a $\mathbf{d} \in \mathcal{D}^* \cap (\mathbf{g} + [0,1)^n)$ with $\mathbf{d} \notin \mathbf{Z}^n$, then again $\mu(T \cap (\mathbf{d} + T)) > 0$, a contradiction. So

$$(3.9) \qquad\qquad \mu(T \cap (\mathbf{g} + [0,1)^n)) > 0 \text{ and } \mathbf{g} + [0,1)^n \nsubseteq T.$$

Notice that $\bigcup_{\mathbf{d} \in \mathcal{D}^*_{B,l}} (T + \mathbf{d})$ covers $\mathbf{g} + [0,1)^n$ (in measure–disjoint fashion). So there exists at least one $\mathbf{0} \neq \mathbf{d} \in \mathcal{D}^*_{B,l}$ such that

$$(3.10)$$
$$\mu((T + \mathbf{d}) \cap (\mathbf{g} + [0,1)^n)) > 0, \text{ or equivalently } \mu(T \cap (\mathbf{g} - \mathbf{d} + [0,1)^n)) > 0.$$

This gives rise to two cases: $\mathbf{d} \in \mathbf{Z}^n$ and $\mathbf{d} \notin \mathbf{Z}^n$.

If $\mathbf{d} \in \mathbf{Z}^n$, then $\mathbf{g} - \mathbf{d} \in \mathbf{Z}^n$ and (3.10) implies $\mathbf{g} - \mathbf{d} \in \mathbf{Z}^n_+$. So $\mathbf{g} - \mathbf{d} \prec \mathbf{g}$ and hence $T \supseteq \mathbf{g} - \mathbf{d} + [0,1)^n$, which together with (3.9) contradicts measure–disjointness.

Now suppose that $\mathbf{d} \notin \mathbf{Z}^n$. Let $\mathbf{d} = \mathbf{d}_0 + \cdots + B^k \mathbf{d}_k$ where all $\mathbf{d}_i \in \mathcal{D}^*$ and $\mathbf{d}_k \neq \mathbf{0}$. If $k \geq 1$ then clearly $\mathbf{d} - \mathbf{d}_i \in \mathbf{Z}^n_+$ and $|\mathbf{d} - \mathbf{d}_i|_1 \geq 1$ for all $\mathbf{d}_i$. So each $\mathbf{d}_i \in \mathbf{g}_i + [0,1)^n$ for some $\mathbf{g}_i \prec \mathbf{g}$ and hence $\mathbf{g}_i \in \mathbf{Z}^n$. But this implies $\mathbf{d} \in \mathbf{Z}^n$, a contradiction. Therefore $k = 0$ and $\mathbf{d} \in \mathcal{D}^*$. Furthermore $\mathbf{d} \in \mathbf{g} + [0,1)^n$ because if not then there would be a $\mathbf{g}' \prec \mathbf{g}$ such that $\mathbf{d} \in \mathbf{g}' + [0,1)^n$, which again would give $\mathbf{d} \in \mathbf{Z}^n$ and hence a contradiction. We now show that this contradicts the measure–disjointness condition. Notice that any $\mathbf{0} \neq \mathbf{d} \in \mathcal{D}^*_{B,l}$ satisfying (3.10) must lie in $\mathbf{g} + [0,1)^n$, so there exists exactly *one* such $\mathbf{0} \neq \mathbf{d} \in \mathcal{D}^*_{B,l}$. Hence

$$T \cup (T + \mathbf{d}) \supseteq \mathbf{g} + [0,1)^n.$$

Let $\mathbf{f}$ be the unique vector in $\mathbf{d} + [0,1)^n \cap \mathbf{Z}^n$. So $\mathbf{f} - \mathbf{g}$ is a zero–one vector, $\mathbf{f} - \mathbf{g} \in \mathcal{D}^*$. Since $[0,1]^n \subseteq T$ it follows that

$$\mu((T + \mathbf{d}) \cap (T + \mathbf{f} - \mathbf{g})) > 0,$$

a contradiction. So we have proved our hypothesis (i) and (ii), and Claim 3 follows.

Finally, we prove that $(\mathcal{E} - \mathcal{E}) \cap (\mathcal{D}^*_{B,\infty} - \mathcal{D}^*_{B,\infty}) = \{\mathbf{0}\}$. Suppose not, then there exist $\mathbf{f}_1, \mathbf{f}_2 \in \mathcal{E}$ and $\mathbf{g}_1, \mathbf{g}_2 \in \mathcal{D}^*_{B,\infty}$ such that $\mathbf{f}_1 - \mathbf{f}_2 = \mathbf{g}_1 - \mathbf{g}_2 \neq \mathbf{0}$. So $\mathbf{f}_1 + \mathbf{g}_2 = \mathbf{f}_2 + \mathbf{g}_1$ and hence

$$\mu((T + \mathbf{g}_1) \cap (T + \mathbf{g}_2)) > 0.$$

This contradicts the measure–disjointness property (2.8).                          $\square$

## 4. Diagonal Case: Zero–One Polynomials

Let $B = \text{diag}(b_1, \ldots, b_n)$ be an expanding nonnegative diagonal matrix. We now show that the criterion of Theorem 3.2 for $(B, \mathcal{D})$ to be nonnegative and minimal feasible can be reformulated in terms of polynomial factorizations of zero–one polynomials. Let $\mathcal{X}$ be a finite subset of $\mathbf{Z}^n_+ := \mathbf{Z}^n \cap \mathbf{R}^n_+$ and assign to it the *generating polynomial*

$$(4.1) \qquad\qquad p_{\mathcal{X}}(\mathbf{z}) = p_{\mathcal{X}}(z_1, \ldots, z_n) := \sum_{\mathbf{d} \in \mathcal{X}} \mathbf{z}^{\mathbf{d}},$$

in which $\mathbf{z}^{\mathbf{d}} := z_1^{d_1} z_2^{d_2} \cdots z_n^{d_n}$. Such a polynomial is just a zero–one polynomial, i.e. $p_{\mathcal{X}}(\mathbf{z}) \in \mathbf{Z}[z_1, \ldots, z_n]$ with all coefficients zero or one.

According to Theorem 3.2 we have

$$(4.2) \qquad\qquad T(B, \mathcal{D}^*) = \bigcup_{\mathbf{f} \in \mathcal{E}} (\mathbf{f} + [0,1]^n)$$

for some finite $\mathcal{E} \subset \mathbf{Z}_+^n$ such that $(\mathcal{E} - \mathcal{E}) \cap (\mathcal{D}_{B,\infty}^* - \mathcal{D}_{B,\infty}^*) = \{\mathbf{0}\}$. Set

$$(4.3) \qquad \mathcal{B} = \left\{ [g_1, \, g_2, \, \cdots, \, g_n]^T \, : \, 0 \le g_i < b_i \right\}.$$

Now the functional equation (3.7) can be encoded using (4.2) as the polynomial function identity

$$(4.4) \qquad p_{\mathcal{E}}(\mathbf{z}^B) p_{\mathcal{B}}(\mathbf{z}) = p_{\mathcal{E}}(\mathbf{z}) p_{\mathcal{D}^*}(\mathbf{z}),$$

where $\mathbf{z}^B := (z_1^{b_1}, \ldots, z_n^{b_n})$.

More generally, one can consider the polynomial factorization identity

$$(4.5) \qquad f(\mathbf{z}^B) g(\mathbf{z}) = f(\mathbf{z}) h(\mathbf{z}),$$

where $f(\mathbf{z})$, $g(\mathbf{z})$, $h(\mathbf{z})$ are all zero–one polynomials.

Classifying all solutions of (4.5) seems an interesting problem. It includes many solutions other than those coming from minimal feasible digit sets. For example, special cases of this identity arise from integer self–affine tiles $T(B, \mathcal{D})$ (as defined in [10]) that are unions of lattice cubes; these include examples for which $(B, \mathcal{D})$ is not feasible, with $[0, 1]^n \not\subseteq T(B, \mathcal{D})$.

Here we will only prove a result that classifies all solutions to (4.5) that satisfy some stringent side conditions, which however cover all cases (4.4). Let $D = \mathrm{diag}(d_1, d_2, \ldots, d_n)$ be a diagonal matrix in which all $d_i$ are integers with $d_i > 1$, and let $f(\mathbf{z}) = \sum_{\mathbf{g} \in \mathcal{X}} a_{\mathbf{g}} \mathbf{z}^{\mathbf{g}}$ be a polynomial. We define

$$(4.6) \qquad [f]_D(\mathbf{z}) := \sum_{\mathbf{g} \in \mathcal{X}} a'_{\mathbf{g}} \mathbf{z}^{\mathbf{g}}$$

where

$$a'_{\mathbf{g}} = \begin{cases} a_{\mathbf{g}}, & \text{if } \mathbf{g} = [g_1, \ldots, g_n]^T \text{ with all } g_i < d_i, \\ 0, & \text{otherwise.} \end{cases}$$

**Theorem 4.1.** *Let* $D = \mathrm{diag}(d_1, \ldots, d_n)$ *in which all* $d_i > 1$ *are integers. Suppose* $f(\mathbf{z})$, $g(\mathbf{z})$, $h(\mathbf{z})$ *are all zero–one polynomials with* $f(\mathbf{0}) = g(\mathbf{0}) = h(\mathbf{0}) = 1$ *satisfying*

$$(4.7) \qquad f(\mathbf{z}^D) g(\mathbf{z}) = f(\mathbf{z}) h(\mathbf{z}).$$

*Suppose further that*

  (i)  *$[g]_D(\mathbf{z}) = g(\mathbf{z})$ and $g(\mathbf{z})$ has no factors of multiplicity greater than 1.*
  (ii)  *For any $\mathbf{f} \in \mathbf{Z}_+^n$, any zero–one vector $\mathbf{d} \ne \mathbf{0}$, and any integer $m \ge 0$, the coefficient of either $\mathbf{z}^{D^m \mathbf{f}}$ or $\mathbf{z}^{D^m(\mathbf{f}+\mathbf{d})}$ is 0 in $f(\mathbf{z})$.*

*Then there exist zero–one polynomials $g_0(\mathbf{z})$, $g_1(\mathbf{z})$, $\ldots$, $g_m(\mathbf{z})$ such that*

$$(4.8) \qquad g(\mathbf{z}) \;=\; g_0(\mathbf{z}) g_1(\mathbf{z}) \cdots g_m(\mathbf{z}),$$

$$(4.9) \qquad h(\mathbf{z}) \;=\; g_0(\mathbf{z}) g_1(\mathbf{z}^D) \cdots g_m(\mathbf{z}^{D^m}),$$

$$(4.10) \qquad f(\mathbf{z}) \;=\; \prod_{i=1}^{m} \prod_{j=0}^{i} g_i(\mathbf{z}^{D^j}).$$

*Remark.* (1). It is possible that some of the $g_i(\mathbf{z}) = 1$. (2). The rather strange–looking hypothesis (ii) actually encodes a non–overlapping property that the set $\mathcal{E}$ in (4.3) possesses.

We will derive this theorem recursively from the following lemma.

**Lemma 4.2.** *Assume that $f(\mathbf{z})$, $g(\mathbf{z})$, $h(\mathbf{z})$ are zero–one polynomials satisfying*

$$(4.11) \qquad\qquad f(\mathbf{z}^D)g(\mathbf{z}) = f(\mathbf{z})h(\mathbf{z})$$

*and all the other hypothesis of Theorem 4.1. Then*

$$(4.12) \qquad\qquad g(\mathbf{z}) = [f]_D(\mathbf{z})[h]_D(\mathbf{z}),$$

*and*

$$(4.13) \qquad\qquad f(\mathbf{z}) = [f]_D(\mathbf{z})f_1(\mathbf{z}^D),$$
$$(4.14) \qquad\qquad h(\mathbf{z}) = [h]_D(\mathbf{z})h_1(\mathbf{z}^D),$$

*where $f_1(\mathbf{z})$ and $h_1(\mathbf{z})$ are zero–one polynomials.*

*Proof.* We write

$$
\begin{aligned}
f(\mathbf{z}) &= a_0(\mathbf{z})\mathbf{z}^{D\mathbf{f}_0} + a_1(\mathbf{z})\mathbf{z}^{D\mathbf{f}_1} + \cdots + a_k(\mathbf{z})\mathbf{z}^{D\mathbf{f}_k}, \\
h(\mathbf{z}) &= b_0(\mathbf{z})\mathbf{z}^{D\mathbf{h}_0} + b_1(\mathbf{z})\mathbf{z}^{D\mathbf{h}_1} + \cdots + b_k(\mathbf{z})\mathbf{z}^{D\mathbf{h}_l},
\end{aligned}
$$

where all $a_i(\mathbf{z})$ and $b_i(\mathbf{z})$ satisfy

$$[a_i]_D(\mathbf{z}) = a_i(\mathbf{z}) \neq 0, \ [b_j]_D(\mathbf{z}) = b_j(\mathbf{z}) \neq 0,$$

and $\mathbf{f}_0 = \mathbf{h}_0 = \mathbf{0}$ and the $\{\mathbf{f}_i\}$ (resp. $\{\mathbf{h}_i\}$) are all distinct nonnegative vectors. Note also that

$$(4.15) \qquad\qquad a_0(\mathbf{z}) = [f]_D(\mathbf{z}), \ b_0(\mathbf{z}) = [g]_D(\mathbf{z}).$$

We first show that

$$a_0(\mathbf{z})b_0(\mathbf{z}) = g(\mathbf{z}),$$

which is (4.12). Since $f(\mathbf{z}) = 1 + \{\text{other terms}\}$, (4.11) yields

$$g(\mathbf{z}) = [f(\mathbf{z}^D)g(\mathbf{z})]_D = [f(\mathbf{z})h(\mathbf{z})]_D = [a_0(\mathbf{z})b_0(\mathbf{z})]_D.$$

Thus if $a_0(\mathbf{z})b_0(\mathbf{z}) \neq g(\mathbf{z})$, then $a_0(\mathbf{z})b_0(\mathbf{z})$ contains some monomial $\mathbf{z}^{\mathbf{g}+D\mathbf{e}}$ having $\mathbf{e} \neq \mathbf{0}$ and $\mathbf{g} = [g_1, \cdots, g_n]^T$ with $g_i < d_i$ for all $i$. Also since $\deg_{z_i}(a_0(\mathbf{z})b_0(\mathbf{z})) \leq 2(d_i - 1)$ the vector $\mathbf{e}$ must be a zero–one vector. Therefore, again from (4.11), since $g(\mathbf{z}) = [g(\mathbf{z})]_D$ by (i), the only monomial in $f(\mathbf{z}^D)$ that can produce such a monomial in $f(\mathbf{z}^D)g(\mathbf{z})$ is $\mathbf{z}^{D\mathbf{e}}$. Thus $\mathbf{z}^{\mathbf{e}}$ must be a monomial in $f(\mathbf{z})$. But $\mathbf{e}$ is a nontrivial zero–one vector, and this contradicts (ii), taking $\mathbf{f} = \mathbf{0}$, $\mathbf{d} = \mathbf{e}$, $m = 0$.

We next prove that for all $1 \leq i \leq k$ and $1 \leq j \leq l$,

$$(4.16) \qquad\qquad a_i(\mathbf{z}) = a_0(\mathbf{z}), \ b_j(\mathbf{z}) = b_0(\mathbf{z}).$$

We arrange the vectors $\{\mathbf{f}_i, \mathbf{h}_j : 0 \leq i \leq k, 0 \leq j \leq l\}$ into a sequence $\{\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{k+l+2}\}$ in such a way that if $|\mathbf{t}_i|_1 < |\mathbf{t}_j|_1$ then we necessarily have $i < j$. (Here $|\mathbf{t}|_1$ denotes the sum of the coordinates of $\mathbf{t}$, its $l^1$–norm.) We prove (4.16) by establishing the following hypothesis, by induction on $m$: *if $\mathbf{t}_m = \mathbf{f}_i$ (resp. $\mathbf{t}_m = \mathbf{h}_j$) then $a_i(\mathbf{z}) = a_0(\mathbf{z})$ (resp. $b_j(\mathbf{z}) = b_0(\mathbf{z})$).*

The hypothesis is clearly true for $m = 1$, since $\mathbf{t}_1$ is either $\mathbf{f}_0$ or $\mathbf{h}_0$.

Suppose that the hypothesis is true for $m' < m$. Now, $\mathbf{t}_m = \mathbf{f}_i$ for some $0 < i \leq k$ or $\mathbf{t}_m = \mathbf{h}_j$ for some $0 < j \leq l$. If $\mathbf{t}_m = \mathbf{f}_i$, then we consider the term $a_i(\mathbf{z})b_0(\mathbf{z})\mathbf{z}^{D\mathbf{f}_i}$ in the expansion of $f(\mathbf{z})h(\mathbf{z})$. We first observe that $a_i(\mathbf{z})b_0(\mathbf{z})$ must be a zero–one polynomial because $f(\mathbf{z})h(\mathbf{z}) = f(\mathbf{z}^D)g(\mathbf{z})$ is. Next we observe that because $b_0(\mathbf{0}) = 1$, all terms in $a_i(\mathbf{z})\mathbf{z}^{D\mathbf{f}_i}$ are also terms in $f(\mathbf{z})h(\mathbf{z}) = f(\mathbf{z}^D)g(\mathbf{z})$. Hence $\mathbf{z}^{D\mathbf{f}_i}$ is a term in $f(\mathbf{z}^D)$ and so $\mathbf{z}^{\mathbf{f}_i}$ is a term in $f(\mathbf{z})$. We claim that $g(\mathbf{z}) - a_i(\mathbf{z})b_0(\mathbf{z})$ must also be a zero–one polynomial. If not, then from (4.11) there is a term $\mathbf{z}^{\mathbf{g}+D\mathbf{f}}$ in $a_i(\mathbf{z})b_0(\mathbf{z})$, where $\mathbf{f} \neq \mathbf{0}$ is a zero–one vector while $\mathbf{g} = [g_1, \cdots, g_n]^T$ satisfy

$g_i < d_i$ for all $i$. It follows from (4.11) that $\mathbf{z}^{D(\mathbf{f}+\mathbf{f}_i)}$ is a term in $f(\mathbf{z}^D)$, so $\mathbf{z}^{\mathbf{f}+\mathbf{f}_i}$ is a term in $f(\mathbf{z})$. But $\mathbf{z}^{\mathbf{f}_i}$ is also a term in $f(\mathbf{z})$. This is a contradiction. Therefore, both $a_i(\mathbf{z})b_0(\mathbf{z})$ and $g(\mathbf{z}) - a_i(\mathbf{z})b_0(\mathbf{z})$ are zero–one polynomials.

Now suppose that $a_i(\mathbf{z}) \neq a_0(\mathbf{z})$. Then $g(\mathbf{z}) - a_i(\mathbf{z})b_0(\mathbf{z}) \neq 0$. So in order for (4.11) to hold there must be some $i' \neq i$, $j' \neq 0$ such that $D\mathbf{f}_{i'} + D\mathbf{h}_{j'} = D\mathbf{f}_i$, $\mathbf{f}_{i'} + \mathbf{h}_{j'} = \mathbf{f}_i$. If $i' > 0$ and $j' > 0$, then both $\mathbf{f}_{i'}$ and $\mathbf{h}_{j'}$ are ahead of $\mathbf{f}_i$ in the sequence $\{\mathbf{t}_j\}$. So $a_{i'}(\mathbf{z}) = a_0(\mathbf{z})$ and $b_{j'}(\mathbf{z}) = b_0(\mathbf{z})$. But this would give $a_{i'}(\mathbf{z})b_{j'}(\mathbf{z}) = g(\mathbf{z})$ and hence $a_i(\mathbf{z})b_0(\mathbf{z}) = 0$, which is a contradiction. Therefore $i' = 0$ (since $j' \neq 0$), and so $\mathbf{f}_0 + \mathbf{h}_{j'} = \mathbf{h}_{j'} = \mathbf{f}_i$. It follows that

$$g(\mathbf{z}) = a_i(\mathbf{z})b_0(\mathbf{z}) + a_0(\mathbf{z})b_{j'}(\mathbf{z}).$$

But $g(\mathbf{z}) = a_0(\mathbf{z})b_0(\mathbf{z})$ and $g(\mathbf{z})$ has no factors with multiplicity greater than 1, so $a_0(\mathbf{z})$ and $b_0(\mathbf{z})$ must be relatively prime. On the other hand,

$$a_i(\mathbf{z})b_0(\mathbf{z}) = g(\mathbf{z}) - a_0(\mathbf{z})b_{j'}(\mathbf{z}) = a_0(\mathbf{z})\big(b_0(\mathbf{z}) - b_{j'}(\mathbf{z})\big).$$

Hence $a_0(\mathbf{z})|a_i(\mathbf{z})$, so by the nonnegativity $b_0(\mathbf{z})a_i(\mathbf{z}) = g(\mathbf{z})$ and thus $a_i(\mathbf{z}) = a_0(\mathbf{z})$, contradicting our assumption.

In the case of $\mathbf{t}_m = \mathbf{h}_j$, $b_j(\mathbf{z}) = b_0(\mathbf{z})$ is proved in the identical fashion. Thus (4.16) is proved.

Finally (4.15) and (4.16) combine to prove (4.13) and (4.14).     $\square$

*Proof of Theorem 4.1.* Let $g_0(\mathbf{z}) = [h]_D(\mathbf{z})$ and $g^*(\mathbf{z}) = [f]_D(\mathbf{z}) = g(\mathbf{z})/g_0(\mathbf{z})$. Then from Lemma 4.2 we have

$$g^*(\mathbf{z}^D)f_1(\mathbf{z}^{D^2})g_0(\mathbf{z})g^*(\mathbf{z}) = g_0(\mathbf{z})h_1(\mathbf{z}^D)g^*(\mathbf{z})f_1(\mathbf{z}^D).$$

This gives $g^*(\mathbf{z}^D)f_1(\mathbf{z}^{D^2}) = h_1(\mathbf{z}^D)f_1(\mathbf{z}^D)$ and hence

$$(4.17) \qquad f_1(\mathbf{z}^D)g^*(\mathbf{z}) = h_1(\mathbf{z})f_1(\mathbf{z}).$$

It is clear that $f_1(\mathbf{z})$, $g^*(\mathbf{z})$, $h_1(\mathbf{z})$ also satisfy the assumptions of Lemma 4.2. Hence we may repeat the above process by setting $g_1(\mathbf{z}) = [h_1]_D(\mathbf{z})$ etc. This process eventually gives us (4.8)–(4.10).     $\square$

We now go back to our digit sets $\mathcal{E}$, $\mathcal{B}$, and $\mathcal{D}^*$, which satisfy (4.4). Now all the hypotheses of Theorem 4.1 are satisfied for $f(\mathbf{z})$, $g(\mathbf{z})$, $h(\mathbf{z})$ equal to $p_{\mathcal{E}}(\mathbf{z})$, $p_{\mathcal{B}}(\mathbf{z})$, $p_{\mathcal{D}^*}(\mathbf{z})$, respectively. For (i) clearly holds for $p_{\mathcal{E}}(\mathbf{z})$ and property (ii) holds as a consequence of Theorem 3.2: $\mathcal{D}^*$ contains all zero–one vectors and $(\mathcal{E}-\mathcal{E}) \cap (\mathcal{D}^*_{B,\infty} - \mathcal{D}^*_{B,\infty}) = \{\mathbf{0}\}$. Thus Theorem 4.1 shows there exist finite subsets $\mathcal{B}_0, \mathcal{B}_1, \ldots, \mathcal{B}_m$ of $\mathbf{Z}_+^n$ such that

$$(4.18) \qquad p_{\mathcal{B}}(\mathbf{z}) = p_{\mathcal{B}_0}(\mathbf{z})p_{\mathcal{B}_1}(\mathbf{z}) \cdots p_{\mathcal{B}_m}(\mathbf{z}),$$

$$(4.19) \qquad p_{\mathcal{D}^*}(\mathbf{z}) = p_{\mathcal{B}_0}(\mathbf{z})p_{\mathcal{B}_1}(\mathbf{z}^B) \cdots p_{\mathcal{B}_m}(\mathbf{z}^{B^m}),$$

$$(4.20) \qquad p_{\mathcal{E}}(\mathbf{z}) = \prod_{i=1}^{m}\prod_{j=0}^{i} p_{\mathcal{B}_i}(\mathbf{z}^{B^j}).$$

Using these formulae, we classify all nonnegative and minimal feasible pairs $(B, \mathcal{D})$ when $B$ is a diagonal matrix.

**Theorem 4.3.** *Suppose that $B = \mathrm{diag}(b_1, b_2, \ldots, b_n)$, with all $b_i > 1$ and $\mathcal{D} \subset \mathbf{R}_+^n$. Then $(B, \mathcal{D})$ is minimal feasible if and only if $b_i \in \mathbf{Z}$ for all $1 \leq i \leq n$ and $\mathcal{D}$ is the Cartesian product of one–dimensional digit sets $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n$, i.e.*

$$(4.21) \qquad\qquad \mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2 \times \cdots \times \mathcal{D}_n,$$

*where* $(b_i, \mathcal{D}_i)$ *are minimal feasible for all* $1 \leq i \leq n$.

*Proof.* We first prove the "only if" part of the theorem. Suppose that $(B, \mathcal{D})$ is nonnegative and minimal feasible. Then by Theorem 3.2 there exists a unique diagonal matrix $S = \text{diag}(s_1, s_2, \ldots, s_n)$, $s_i > 0$, such that $\mathcal{D}^* = S(\mathcal{D}) \subset \mathbf{Z}_+^n$ contains all zero–one vectors, and by the discussion above (4.18)–(4.20) hold. Now

$$\mathcal{B} = \left\{ [g_1, g_2, \ldots, g_n]^T \, : \, 0 \leq g_i < b_i \text{ for all } i \right\},$$

so

$$(4.22) \qquad\qquad p_{\mathcal{B}}(\mathbf{z}) = \prod_{i=1}^{n} q_i(z_i)$$

where $q_i(\mathbf{z}) = 1 + z_i + z_i^2 + \cdots + z_i^{b_i - 1}$ is a polynomial in $z_i$ alone. Now set

$$\varphi_{i,j}(z_i) := \text{g.c.d.}(q_i(z_i), p_{\mathcal{B}_j}(\mathbf{z})),$$

noting that $\varphi_{i,j}(z_i)$ divides $q_i(z_i)$ so is a polynomial in $z_i$ alone. It is determined up to a multiplicative constant factor, and since $p_{\mathcal{B}}(\mathbf{0}) = 1$, (4.18) shows that all $p_{\mathcal{B}_j}(\mathbf{0}) = 1$, hence $\varphi_{i,j}(0) \neq 0$ so we may normalize it by requiring $\varphi_{i,j}(0) = 1$. Now (4.18) and (4.22) together imply that

$$(4.23) \qquad\qquad p_{\mathcal{B}_j}(\mathbf{z}) = c_0 \prod_{i=1}^{n} \varphi_{i,j}(z_i)$$

for some constants $c_0$, and taking $\mathbf{z} = \mathbf{0}$ shows $c_0 = 1$. Now (4.23) implies that each $\varphi_{i,j}(z_i)$ is a zero–one polynomial, because $p_{\mathcal{B}_j}(\mathbf{z})$ is, and each of the $\varphi_{i,j}(z_i)$ depends on $z_i$ only. We now substitute the expression (4.23) into (4.19) and obtain

$$(4.24) \qquad\qquad p_{\mathcal{D}^*}(\mathbf{z}) = \prod_{j=1}^{m} \prod_{i=1}^{n} \varphi_{i,j}(z_i^{b_i^j}) = \prod_{i=1}^{n} \Big( \prod_{j=1}^{m} \varphi_{i,j}(z_i^{b_i^j}) \Big).$$

Next we observe that $\prod_{j=1}^{m} \varphi_{i,j}(z_i^{b_i^j})$ is a zero–one polynomial of $z_i$, for $\varphi_{i,j}(z_i)$ is a zero–one polynomial of degree at most $b_i - 1$; it therefore is a generating polynomial $p_{\mathcal{D}_i^*}(z_i)$ for some subset $\mathcal{D}_i^* \subset \mathbf{Z}$. Now (4.19) becomes

$$(4.25) \qquad\qquad p_{\mathcal{D}^*}(\mathbf{z}) = \prod_{i=1}^{n} p_{\mathcal{D}_i^*}(z_i)$$

and hence

$$\mathcal{D}^* = \mathcal{D}_1^* \times \mathcal{D}_2^* \times \cdots \times \mathcal{D}_n^*.$$

Thus

$$\mathcal{D} = S^{-1}(\mathcal{D}^*) = \mathcal{D}_1 \times \mathcal{D}_2 \times \cdots \times \mathcal{D}_n$$

where $\mathcal{D}_i = s_i^{-1} \mathcal{D}_i^*$, and each $(b_i, \mathcal{D}_i)$ is nonnegative. Notice that

$$T(B, \mathcal{D}) = T(b_1, \mathcal{D}_1) \times T(b_1, \mathcal{D}_2) \times \cdots \times T(b_1, \mathcal{D}_n),$$

and $T(B, \mathcal{D})$ has Lebesgue measure equal to the product of the one–dimensional Lebesgue measures of $T(b_i, \mathcal{D}_i)$. Thus the feasibility of $(B, \mathcal{D})$ implies that of each $(b_i, \mathcal{D}_i)$, whence $|\mathcal{D}_i| \geq b_i$. However

$$|\mathcal{D}_1| \cdots |\mathcal{D}_n| = |\det(B)| = b_1 \cdots b_n,$$

and it follows that $|\mathcal{D}_i| = b_i$, so each $(b_i, \mathcal{D}_i)$ is minimal feasible.

   We now prove the "if" part of the theorem. If $(b_i, \mathcal{D}_i)$ is nonnegative and minimal feasible for all $1 \leq i \leq n$, then clearly every $\mathbf{x} \in \mathbf{R}_+^n$ has a radix expansion using

base $B$ and digits from $\mathcal{D} = \mathcal{D}_1 \times \mathcal{D}_2 \times \cdots \times \mathcal{D}_n$. Hence $(B, \mathcal{D})$ must be feasible, and nonnegativity and minimality are obvious.     $\square$

## 5. GENERAL CASE

The results of Section 2 showed that if $(A, \mathcal{D})$ is nonnegative and feasible then $A = BP$ where $B$ is a diagonal matrix and $P$ is a permutation matrix. We now classify all nonnegative minimal feasible pairs $(A, \mathcal{D})$ for such $A$.

For any permutation $\sigma$ of $\{1, 2, \ldots, n\}$, let $P_\sigma$ be the permutation matrix

$$(P_\sigma)_{i,j} = \begin{cases} 1 & \text{if } j = \sigma(i), \\ 0 & \text{otherwise.} \end{cases}$$

Consider the cyclic decomposition of $\sigma$, $\sigma = \tau_1 \tau_2 \cdots \tau_k$ into disjoint cycles, e.g.

$$(1, 3, 4)(2, 5)(6) \text{ represents } \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 3 & 5 & 4 & 1 & 2 & 6 \end{pmatrix}.$$

For any $i \in \{1, 2, \ldots, n\}$ write $i \in \tau_j = (j_1, \ldots, j_l)$ if some $j_m = i$.

Let $A = BP_\sigma$ where $B = \text{diag}(b_1, \ldots b_n)$. Now $A$ splits up into blocks corresponding to the cycle decomposition $\sigma = (\tau_1)(\tau_2) \cdots (\tau_k)$ of $\sigma$. Set

$$(5.1) \qquad\qquad B_{\tau_i} = \text{diag}(b_1^{(i)}, b_2^{(i)}, \ldots, b_n^{(i)})$$

where $b_j^{(i)} = b_j$ if $j \in \tau_i$ and $b_j^{(i)} = 1$ otherwise. Then

$$(5.2) \qquad\qquad A = BP_\sigma = (B_{\tau_1} P_{\tau_1})(B_{\tau_2} P_{\tau_2}) \cdots (B_{\tau_n} P_{\tau_n}),$$

and the matrices $B_{\tau_i} P_{\tau_i}$ all commute pairwise. For example, if $\lambda > 0$ is arbitrary, then

$(5.3)$

$$A = \begin{bmatrix} 0 & 0 & \lambda \\ 0 & 3 & 0 \\ \frac{4}{\lambda} & 0 & 0 \end{bmatrix} = \text{diag}(\lambda, 3, \tfrac{4}{\lambda})P_{(1,3)}P_{(2)} = B_{(1,3)}P_{(1,3)}B_{(2)}P_{(2)},$$

where $B_{(1,3)} = \text{diag}(\lambda, 1, \frac{4}{\lambda})$ and $B_{(2)} = \text{diag}(1, 3, 1)$.

**Lemma 5.1.** *Let $A = BP_\tau$ where $\tau$ is an $n$–cycle, i.e. a cycle of length $n$, and $B = \text{diag}(b_1, \ldots, b_n)$ with $b = b_1 b_2 \cdots b_n \in \mathbf{Z}$. Suppose that $(A, \mathcal{D})$ is nonnegative and minimal feasible. Then there exists an integer $1 \leq k \leq n$ such that all digits $\mathbf{d} \in \mathcal{D}$ lie on the $x_k$–axis, i.e. $\mathbf{d} = d\mathbf{e_k}$ for some $d \in \mathbf{R}_+$ for all $\mathbf{d} \in \mathcal{D}$.*

*Proof.* We first observe that $A^n = bI$ and since $(bI, \mathcal{D}_{A,n})$ is also nonnegative and minimal feasible, by Theorem 4.3 the set

$$\mathcal{D}_{A,n} = \mathcal{D} + A(\mathcal{D}) + \cdots + A^{n-1}(\mathcal{D})$$

must be the Cartesian product of $n$ one–dimensional minimal feasible digit sets. So there must be exactly $n(b - 1)$ nonzero digits in $\mathcal{D}_{A,n}$ that lie on one of the coordinate axes. Since the action of $A$ is to permute coordinate axes, each of these digits has the form $A^k \mathbf{d}$ for a digit $\mathbf{d} \in \mathcal{D}$ lying on some coordinate axis and $0 \leq k \leq n - 1$. Now $\mathbf{0} \in \mathcal{D}$ so there are $b - 1$ nonzero digits $\mathbf{d} \in \mathcal{D}$. Hence we conclude that *all* digits $\mathbf{d} \in \mathcal{D}$ must lie on the $n$ coordinate axes.

It remains to show that all $\mathbf{d} \in \mathcal{D}$ must lie on the *same* coordinate axis, i.e. all $\mathbf{d}_i = d_i \mathbf{e}_l$ for some coordinate axis vector $\mathbf{e}_l$ and some $d_i \in \mathbf{R}_+$. We argue by contradiction. Suppose not, then for each $0 \leq k < n$ the set $A^k(\mathcal{D})$ includes vectors

on at least two different coordinate axes. Now choose a $\mathbf{d}^* \in A^k(\mathcal{D})$ with $0 \le k < n$ such that

$$(5.4) \qquad |\mathbf{d}^*|_1 = \min\Big\{ |\mathbf{d}|_1 \; : \; \mathbf{d} \in \bigcup_{j=0}^{n-1} A^j(\mathcal{D}) \Big\}.$$

Suppose that $\mathbf{d}^* = d^* \mathbf{e}_m$ and choose any $\mathbf{d}' = d' \mathbf{e}_l \in A^k(\mathcal{D})$ such that $l \neq m$. Both $\mathbf{d}^*, \mathbf{d}' \in \mathcal{D}_{A,n}$ and $\mathbf{d}^*$ is orthogonal to $\mathbf{d}'$, while Theorem 4.3 shows that $\mathcal{D}_{A,n}$ is a Cartesian product of coordinate axes, hence $\mathbf{d}^* + \mathbf{d}' \in \mathcal{D}_{A,n}$. Thus there exists a representation

$$(5.5) \qquad d^* \mathbf{e}_m + d' \mathbf{e}_l = \sum_{j=0}^{n-1} A^j \mathbf{d}_j, \text{ all } \mathbf{d}_j \in \mathcal{D}.$$

Since each $A^j \mathbf{d}_j$ lie on a coordinate axis, (5.4) implies that one of the terms on the right–hand side of (5.5) must be $\mathbf{d}^*$. It is necessarily $A^k \mathbf{d}_k = \mathbf{d}^*$, hence

$$d' \mathbf{e}_l = \sum_{0 \le j < n, j \neq k} A^j \mathbf{d}_j, \text{ all } \mathbf{d}_j \in \mathcal{D}.$$

This contradicts the fact that all $b^n$ expansions in $\mathcal{D}_{A,n}$ must be distinct. Thus all $\mathbf{d} \in \mathcal{D}$ must lie on the same coordinate axis. $\qquad\square$

At last we can obtain the desired classification.

**Theorem 5.2.** *Let $A$ be an expanding nonnegative matrix and $\mathcal{D} \subset \mathbf{R}_+^n$. Then $(A, \mathcal{D})$ is minimal feasible if and only if the following conditions hold:*

(i) $A = BP_\sigma$ *where* $B = \mathrm{diag}(b_1, b_2, \ldots, b_n)$ *and* $P_\sigma$ *is a permutation matrix. If $\sigma$ has the cyclic decomposition $\sigma = \tau_1 \tau_2 \cdots \tau_k$, then*

$$(5.6) \qquad b_j^* := \prod_{i \in \tau_j} b_i \in \mathbf{Z}_+, \; 1 \le j \le k,$$

*with each $b_j^* \ge 2$.*

(ii) *There exist one–dimensional digit sets $\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_k \subset \mathbf{R}_+$, with $|\mathcal{D}_j| = b_j^*$, such that $(b_j^*, \mathcal{D}_j)$ are nonnegative and minimal feasible for all $1 \le j \le k$, and*

$$(5.7) \qquad \mathcal{D} = \mathcal{D}_1 \mathbf{e}_{l_1} + \mathcal{D}_2 \mathbf{e}_{l_2} + \cdots + \mathcal{D}_k \mathbf{e}_{l_k}$$

*for coordinate vectors $\mathbf{e}_{l_j}$ with $l_j \in \tau_j$ for all $1 \le j \le k$.*

*Remark.* This theorem immediately implies Theorem 1.2, since for every $A$ satisfying (i), there clearly exists a set $\mathcal{D}$ satisfying (ii).

*Proof.* We first prove the "only if" part. Suppose that $(A, \mathcal{D})$ is nonnegative and minimal feasible. By Lemma 2.3

$$A = BP_\sigma = BP_{\tau_1} P_{\tau_2} \cdots P_{\tau_k}.$$

For each $1 \le j \le k$, set

$$\mathbf{R}^{\tau_j} := \Big\{ [x_1, x_2, \ldots, x_n]^T \; : \; x_i = 0 \text{ if } i \notin \tau_j \Big\}$$

and $\mathbf{R}_+^{\tau_j} := \mathbf{R}^{\tau_j} \cap \mathbf{R}_+^n$. Next set $\mathcal{D}^{\tau_j} := \mathcal{D} \cap \mathbf{R}^{\tau_j}$. Consider the decomposition (5.2) of $A$:

$$A = BP_\sigma = (B_{\tau_1} P_{\tau_1})(B_{\tau_2} P_{\tau_2}) \cdots (B_{\tau_k} P_{\tau_k}).$$

It follows from the nonnegativity of $(A, \mathcal{D})$ and $A(\mathbf{R}^{\tau_j}) = \mathbf{R}^{\tau_j}$ that each $\mathbf{x} \in \mathbf{R}_+^{\tau_j}$ has a radix expansion using only digits in $\mathcal{D}^{\tau_j}$. Notice that for any $\mathbf{d} \in \mathcal{D}^{\tau_j}$ one has

$$A\mathbf{d} = (B_{\tau_j} P_{\tau_j})\mathbf{d}.$$

Thus $(B_{\tau_j} P_{\tau_j}, \mathcal{D}^{\tau_j})$ when viewed as acting on $\mathbf{R}^{\tau_j}$ is a feasible pair, hence

$$(5.8) \qquad c_j := |\mathcal{D}^{\tau_j}| \geq |\det(B_{\tau_j} P_{\tau_j})| = b_j^*.$$

Let $m \geq 1$ such that $A^m$ is a diagonal matrix, and recall the pair $(A^m, \mathcal{D}_{A,m})$ is minimal feasible using Lemma 2.1. Clearly

$$\mathcal{D}_{A,m} \cap \mathbf{R}^{\tau_j} = \mathcal{D}^{\tau_j}_{B_{\tau_j} P_{\tau_j}, m},$$

hence

$$(5.9) \qquad |\mathcal{D}_{A,m} \cap \mathbf{R}^{\tau_j}| = |\mathcal{D}^{\tau_j}_{B_{\tau_j} P_{\tau_j}, m}| = |\mathcal{D}^{\tau_j}_{A,m}| = c_j^m.$$

But because $A^m$ is diagonal, it follows from Theorem 4.3 that its digit set $D_{A,m}$ is a direct sum of one–dimensional digit sets, hence in particular

$$(5.10)$$
$$\mathcal{D}_{A,m} \supseteq \mathcal{D}^{\tau_1}_{A,m} + \cdots + \mathcal{D}^{\tau_k}_{A,m} = \mathcal{D}^{\tau_1}_{B_{\tau_1} P_{\tau_1}, m} + \cdots + \mathcal{D}^{\tau_k}_{B_{\tau_k} P_{\tau_k}, m},$$

which is a direct sum because all summands lie in mutually orthogonal subspaces. This gives

$$|\mathcal{D}_{A,m}| = \left(b_1^* b_2^* \cdots b_k^*\right)^m \geq c_1^m c_2^m \cdots c_k^m,$$

using (5.9). Comparison with (5.8) then implies that

$$(5.11) \qquad b_j^* = c_j \in \mathbf{Z}, \ 1 \leq i \leq k,$$

and clearly $b_j^* > 1$. Moreover, it now follows from (5.10) and (5.11) that

$$(5.12) \qquad \mathcal{D}_{A,m} = \mathcal{D}^{\tau_1}_{B_{\tau_1} P_{\tau_1}, m} + \cdots + \mathcal{D}^{\tau_k}_{B_{\tau_k} P_{\tau_k}, m}.$$

Now each $(B_{\tau_j} P_{\tau_j}, \mathcal{D}^{\tau_j})$ viewed as acting on $\mathbf{R}^{\tau_j}$ is a nonnegative minimal feasible pair, using (5.11), and the hypothesis of Lemma 5.1 are satisfied for it. Hence there exists an $l_j \in \tau_j$ such that $\mathcal{D}^{\tau_j} = \mathcal{D}_j \mathbf{e}_{l_j}$ where $\mathcal{D}_j \subset \mathbf{R}$ and each $(b_j^*, \mathcal{D}_j)$ is a nonnegative minimal feasible one–dimensional pair.

To establish (ii) it remains to show that (5.7) holds, i.e. $\mathcal{D} = \tilde{\mathcal{D}}$ where

$$\tilde{\mathcal{D}} = \mathcal{D}^{\tau_1} + \mathcal{D}^{\tau_2} + \cdots + \mathcal{D}^{\tau_k}.$$

To prove this it suffices to show that $\mathcal{D} \subseteq \tilde{\mathcal{D}}$, because they have the same cardinality $b = b_1^* \cdots b_k^*$. We consider the orthogonal projection operator $\pi_i : \mathbf{R}^n \longrightarrow \mathbf{R}^{\tau_i}$ and examine the set $\pi_i(\mathcal{D})$, with elements counted with multiplicity. If we show that all elements of $\pi_i(\mathcal{D})$ lie in $\mathcal{D}^{\tau_i} = \mathcal{D}_i \mathbf{e}_i$ for $1 \leq i \leq k$, then $\mathcal{D} \subseteq \tilde{\mathcal{D}}$. We will actually show the stronger result:

$$(5.13) \qquad \pi_i(\mathcal{D}) \text{ is } \mathcal{D}^{\tau_i} \text{ with all elements of multiplicity } b/b_i^*.$$

Now $\mathbf{R}^{\tau_i}$ is an invariant subspace of $A$, and $\pi_i$ commutes with $B_{\tau_i}$ and $P_{\tau_i}$. In particular, using (5.12) and ignoring multiplicity,

$$\begin{aligned}
\left(\pi_i(\mathcal{D})\right)_{A,m} &= \pi_i(\mathcal{D}_{A,m}) \\
&= \pi_i\left(\mathcal{D}^{\tau_1}_{A,m} + \mathcal{D}^{\tau_2}_{A,m} + \cdots + \mathcal{D}^{\tau_k}_{A,m}\right) \\
&= \mathcal{D}^{\tau_i}_{A,m} \\
(5.14) \qquad &= \left\{ \sum_{j=0}^{m-1} A^j \mathbf{d}_j^{(i)} : \text{ all } \mathbf{d}_j^{(i)} \in \mathcal{D}^{\tau_i} \right\}.
\end{aligned}$$

If we count multiplicity, then (5.12) implies that

$$(5.15) \qquad \left(\pi_i(\mathcal{D})\right)_{A,m} := \pi_i(\mathcal{D}) + A\pi_i(\mathcal{D}) + \cdots + A^{m-1}\pi_i(\mathcal{D})$$

is just $\mathcal{D}_{A,m}^{\tau_i}$ with all elements counted with multiplicity $(b/b_i^*)^m$.

Now the first part of the proof of Lemma 5.1 applies to the set (5.15) to conclude that every element of $\pi_i(\mathcal{D})$ lies on a coordinate axis.

To proceed, we determine the elements of $\mathcal{D}_{A,m}^{\tau_i}$ that lie on some coordinate axis. Now $\tau_i$ is a cycle of order $q$, say $\tau_i = (r_1, r_2, \ldots, r_q)$ with $r_1 = l_i$, say, and necessarily $q$ divides $m$. The decomposition (5.14) factorizes as

$$\sum_{j=0}^{q-1} A^j \left( \sum_{l=0}^{m/q} A^{ql} \mathbf{d}_{ql+j}^{(i)} \right),$$

where the inner sum involves only coordinate vectors along the axis $\mathbf{e}_{r_1} = \mathbf{e}_{l_i}$. In particular $(\pi_i(\mathcal{D}))_{A,m}$ contains vectors on coordinate axis $\mathbf{e}_{r_j}$ with coefficients lying in the sets

(5.16)     $$S_j := \left\{ a_j \left( \mathcal{D}_i + \mathcal{D}_i b_i^* + \cdots + \mathcal{D}_i (b_i^*)^{\frac{m}{k}-1} \right) \right\}$$

where $a_1 = 1$ and $a_j = b_{r_1} b_{r_2} \cdots b_{r_{j-1}}$ for $1 < j < q-1$, and $b_i^* = b_{r_1} \cdots b_{r_q} \geq 2$ is an integer.

Suppose now that $\pi_i(\mathcal{D})$ contains some elements $\mathbf{d} \neq d\mathbf{e}_{r_1}$ for any $d \in \mathbf{R}$. It then lies on some coordinate axis $\mathbf{e}_{r_j}$, and necessarily has a coefficient from the set (5.16). In particular any of its powers $A^j \mathbf{d}$ that lie on the axis $\mathbf{e}_{r_1}$ have coefficients in the set

$$\mathcal{D}_i b_i^* + \cdots + \mathcal{D}_i (b_i^*)^{\frac{m}{k}-1}.$$

Hence they can never appear in the expansion of any elements in $(\pi_i(\mathcal{D}))_{A,m}$ of the form $d'\mathbf{e}_{r_0}$ with $d' \in \mathcal{D}_i$. Thus all representations of such vectors in the expression (5.14) must be of the form

(5.17)     $$d'\mathbf{e}_{r_0} = d'\mathbf{e}_{r_0} + A(\mathbf{0}) + \cdots + A^{k-1}(\mathbf{0}).$$

However $\mathbf{0}$ must occur with its correct multiplicity $b/b_i^*$ in $\pi_i(\mathcal{D})$, in order to get exactly $(b/b_i^*)^m$ copies of $\mathbf{0}$ in $(\pi_i(\mathcal{D}))_{A,m}$, since all terms in (5.14) are nonnegative. Thus the number of representations of $d'\mathbf{e}_{r_0}$ on the right side of (5.17) is

$$(b/b_i^*)^{m-1} \cdot \left( \text{multiplicity of } d'\mathbf{e}_{r_0} \text{ in } \pi_i(\mathcal{D}) \right).$$

This number is $(b/b_i^*)^m$, hence $d'\mathbf{e}_{r_0}$ has multiplicity $b/b_i^*$ in $\pi_i(\mathcal{D})$ for all $d' \in \mathcal{D}_i$. This count exhausts $\pi_i(\mathcal{D})$, and proves (5.13). Hence $\mathcal{D} = \tilde{\mathcal{D}}$, proving (ii).

We now finish the proof by proving the "if" part of the theorem. According to Lemma 5.1, if we view $B_{\tau_j} P_{\tau_j}$ and $\mathcal{D}^{\tau_j} = \mathcal{D}_j \mathbf{e}_{l_j}$ as lying on $\mathbf{R}^{\tau_j}$, then $(B_{\tau_j} P_{\tau_j}, \mathcal{D}^{\tau_j})$ are nonnegative and minimal feasible, and

$$T(A, \mathcal{D}) = T(B_{\tau_1} P_{\tau_1}, \mathcal{D}^{\tau_1}) \times \cdots \times T(B_{\tau_k} P_{\tau_k}, \mathcal{D}^{\tau_k})$$

if the coordinates are suitably numbered. Hence $(A, \mathcal{D})$ is feasible, and nonnegativity and minimality are clear.     $\square$

**Example 5.1.** The matrix

$$A := \begin{bmatrix} 0 & 0 & \lambda \\ 0 & 3 & 0 \\ \frac{4}{\lambda} & 0 & 0 \end{bmatrix} := BP_\sigma := \begin{bmatrix} \lambda & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & \frac{4}{\lambda} \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

given in (5.3) satisfies condition (i) of Theorem 5.2 for all $\lambda > 0$. By Theorem 5.2 $(A, \mathcal{D})$ is feasible for

$$\mathcal{D} = \{0, 1, 8, 9\} \mathbf{e}_1 + \{0, 1, 2\} \mathbf{e}_2$$

because $\{0, 1, 8, 9\}$ is feasible for base $\lambda \cdot \frac{4}{\lambda} = 4$ and $\{0, 1, 2\}$ is feasible for base 3. We can see this more easily from $A^2 = \text{diag}(4, 9, 4)$ and

$$\mathcal{D}_{A,2} = \{0, 1, 8, 9\}\mathbf{e}_1 + \{0, 1, \ldots, 8\}\mathbf{e}_2 + \frac{4}{\lambda} \cdot \{0, 1, 8, 9\}\mathbf{e}_3,$$

which is a direct sum of three one–dimensional feasible digit sets. Note that $B = \text{diag}(\lambda, 3, \frac{4}{\lambda})$ is an expanding matrix only for $1 < \lambda < 4$.

## References

1. C. Bandt, *Self-similar sets 5. integer matrices and fractal tilings of $R^n$*, Proc. Amer. Math Soc. **112** (1991), 549–562. MR **92d**:58093
2. M. Barnsley, *Fractals everywhere*, Academic Press, 1988. MR **90e**:58080
3. L. Carlitz and L. Moser, *On some special factorizations of $(1 - X^n)/(1 - X)$*, Canad. Math. Bull. **9** (1966), 421–426. MR **34**:4262
4. W. Gilbert, *Geometry of radix expansions,* in: The Geometry Vein: the Coxeter Festschrift (1981), 129–139. MR **83j**:12001
5. K. Gröchenig and A. Haas, *Self–similar lattice tilings,* J. Fourier Analysis **1** (1994), 131–170.
6. R. Kenyon, *Self-similar tilings,* Ph.D thesis, Princeton University (1990).
7. R. Kenyon, *Self-replicating tilings,* in: Symbolic Dynamics and Applications (P. Walters, ed.) Contemporary Math. vol. 135 (1992), 239–264. MR **94a**:52043
8. D. E. Knuth, *The art of computer programming: volume 2. Seminumerical algorithms,* Addison–Wesley, 1969. (See Chapter 4.1, exercise 20–24.) MR **44**:3531
9. J.C. Lagarias and Y. Wang, *Self-affine tiles in $\mathbf{R}^n$,* Advances in Math., to appear.
10. J.C. Lagarias and Y. Wang, *Integral self-affine tiles in $\mathbf{R}^n$ I. Standard and nonstandard digit sets,* J. London Math. Soc., to appear.
11. D. W. Matula, *Basic digit sets for radix representations,* J. Assoc. Comput. Mech. **4** (1982), 1131–1143. MR **83k**:68017
12. A. M. Odlyzko, *Nonnegative digit sets in positional number systems*, Proc. London Math. Soc. **37** (1978), 213–229. MR **80m**:10004
13. A. Vince, *Replicating Tesselations,* SIAM J. Discrete Math. **6**, no. 3 (1993), 501–521. MR **94e**:52023

AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974
*E-mail address:* jcl@research.att.com

School of Mathematics, Georgia Institute of Technology, Atlanta, Georgia 30332
*E-mail address:* wang@math.gatech.edu