Uniform Distribution Theory **9** (2014), no.1, 27–42

uniform
distribution
theory

A CAUCHY-DAVENPORT THEOREM FOR SEMIGROUPS

Salvatore Tringali

ABSTRACT. We generalize the Davenport transform and use it to prove that, for a (possibly non-commutative) cancellative semigroup $\mathbb{A} = (A, +)$ and non-empty subsets X, Y of A such that the subsemigroup generated by Y is commutative, we have $|X + Y| \geq \min(\omega(Y), |X| + |Y| - 1)$, where

$$\omega(Y) := \sup_{y_0 \in Y \cap A^\times} \inf_{y \in Y \setminus \{y_0\}} |\langle y - y_0 \rangle|.$$

This carries over the Cauchy-Davenport theorem to the broader setting of semigroups, and it implies, on the one hand, a common extension of I. Chowla's and S.S. Pillai's theorems for cyclic groups, and on the other a significant strengthening of another generalization of the same Cauchy-Davenport theorem (to commutative groups), where $\omega(Y)$ in the above is replaced by the infimum of $|S|$ as S ranges over the non-trivial subgroups of $\mathbb{A}$.

Communicated by Georges Grekos

1. Introduction

The present paper deals with the structure theory of *semigroups*. We refer to [B2], [B1, Chapter I, Sections 1-2, 4, and 6], and [Ho, Chapter 1] for all necessary prerequisites as well as for notation and terminology used but not defined here.

Semigroups are a natural framework for huge parts of theories traditionally developed under more "favorable conditions". Not only this can suggest new directions of research and shed light on questions primarily focused on groups,

2010 Mathematics Subject Classification: 05E15, 11B13 (primary); 20E99, 20M10 (secondary).

Keywords: Cauchy-Davenport theorem, Chowla's theorem, Davenport transform, difference sets, groups, Hamidoune-Károlyi theorem, Kemperman's theorem, Pillai's theorem, semigroups, structure theory, sumsets, transformation proofs.

This research was funded from the European Community's 7th Framework Programme (FP7/2007-2013) under Grant Agreement No. 276487 (project ApProCEM), and partly from the ANR Project No. ANR-12-BS01-0011 (project CAESAR).

but in principle it also makes methods and results otherwise restricted to "richer settings" applicable to larger classes of problems.

Here, a semigroup is a pair $\mathbb{A} = (A, +)$ consisting of a (possibly empty) set A, referred to as the carrier of $\mathbb{A}$, and an associative binary operation $+$ on A (unless otherwise specified, all semigroups considered below are written additively, but they are not necessarily commutative).

Given subsets X, Y of A, we define as usual the sumset, relative to $\mathbb{A}$, of the pair (X, Y) as the set $X+Y := \{x+y : x \in X, y \in Y\}$, which is written as $x+Y$ (respectively, $X+y$) if $X = \{x\}$ (respectively, $Y = \{y\}$). Further, we extend the notion of difference set from groups to semigroups by

$$X - Y := \{z \in A : (z + Y) \cap X \neq \emptyset\}$$

and

$$-X + Y := \{z \in A : (X + z) \cap Y \neq \emptyset\}.$$

Expressions of the form $Z_1 + \cdots + Z_n$ or $\sum_{i=1}^n Z_i$, involving one or more summands, or of the form $-x+Y$ or $X-y$ for $x, y \in A$ are defined in a similar way; in particular, we use nZ for $Z_1 + \cdots + Z_n$ if the Z_i are all equal to the same Z.

We say that $\mathbb{A}$ is unital, or a *monoid*, if there exists $0 \in A$ such that $z + 0 = 0 + z = z$ for all z; when this is the case, 0 is unique and called *the* identity of $\mathbb{A}$. Then, we let $\mathbb{A}^\times$ be the set of units of $\mathbb{A}$, in such a way that $\mathbb{A}^\times := \emptyset$ if $\mathbb{A}$ is not a monoid; this is simply denoted as $A^\times$ if there is no likelihood of confusion. If $\mathbb{A}$ is unital with identity 0, a *unit* of $\mathbb{A}$ is now an element z for which there exists $\tilde{z}$, provably unique and called *the* inverse of z (in $\mathbb{A}$), such that $z + \tilde{z} = \tilde{z} + z = 0$. Moreover, if Z is a subset of A, we write $\langle Z \rangle_{\mathbb{A}}$ for the smallest subsemigroup of $\mathbb{A}$ containing Z, and given $z \in A$ we use $\langle z \rangle_{\mathbb{A}}$ for $\langle \{z\} \rangle_{\mathbb{A}}$ and $\text{ord}_{\mathbb{A}}(z)$ for the order of z (in $\mathbb{A}$); i.e., we set $\text{ord}_{\mathbb{A}}(z) := |\langle z \rangle_{\mathbb{A}}|$, so generalizing the notion of order for the elements of a group. Here and later, the subscript '$\mathbb{A}$' may be omitted if $\mathbb{A}$ is clear from the context. Finally, we say that $\mathbb{A}$ is cancellative if for $x, y, z \in A$ it holds $z + x = z + y$ or $x + z = y + z$ only if $x = y$; of course, any group is a cancellative monoid.

Sumsets in (mostly commutative) groups have been intensively investigated for several years (see [Ru] for a recent survey), and interesting results have been also obtained in the case of commutative cancellative monoids (see [G] and references therein, where these structures are simply called "monoids"). The present paper aims to extend aspects of the theory to the more general setting of possibly *non-commutative* semigroups.

Historically, one of the first significant achievements in the field is the Cauchy-Davenport theorem, originally established by A.-L. Cauchy [C] in 1813, and independently rediscovered by H. Davenport [D1, D2] more than a century later:

Theorem 1 (Cauchy-Davenport theorem). *Let $(A,+)$ be a group of prime order p and X,Y non-empty subsets of A. Then, $|X+Y| \ge \min(p, |X|+|Y|-1)$.*

The result has been the subject of numerous papers, and received many different proofs, each favoring alternative points of view and eventually leading to progress on a number of other questions. In fact, the main contribution here is an extension of Theorem 1 to cancellative semigroups (see Section 2).

The Cauchy-Davenport theorem applies especially to the additive group of the integers modulo a prime. Extensions to composite moduli have been given by several authors, and notably by I. Chowla [Ch] and S.S. Pillai [P]. These results, reported below for the sake of exposition and used by Chowla and Pillai in relation to Waring's problem, are further strengthened, in Section 2, by Corollary 15, which can be viewed as a common generalization of both of them, and whose proof is sensibly shorter than each of the proofs appearing in [Ch] and [P] (not to mention that it comes as a by-product of a deeper result).

In what follows, if m is a positive integer, we write $\mathbb{Z}/m\mathbb{Z}$ for the set of integers modulo m, endowed with its standard additive and multiplicative structure.

Theorem 2 (Chowla's theorem). *Let m be an integer ≥ 1. If X,Y are non-empty subsets of $\mathbb{Z}/m\mathbb{Z}$ such that $0 \in Y$ and $\gcd(m,y)=1$ for each $y \in Y \setminus \{0\}$, then $|X+Y| \ge \min(m, |X|+|Y|-1)$.*

Theorem 3 (Pillai's theorem). *Given an integer $m \ge 1$, pick non-empty subsets X,Y of $\mathbb{Z}/m\mathbb{Z}$. Let δ be the maximum of $\gcd(m, y-y_0)$ for distinct $y, y_0 \in Y$ if $|Y| \ge 2$, and set $\delta := 1$ otherwise. Then, $|X+Y| \ge \min(\delta^{-1}m, |X|+|Y|-1)$.*

A partial account of further results in the same spirit can be found in [N, Section 2.3], along with an entire chapter dedicated to Kneser's theorem [N, Chapter 4], which, among the other things, implies Theorem 2 (and hence Theorem 1); see [N, Section 4.6, Exercises 5 and 6].

Generalizations of the Cauchy-Davenport theorem of a somewhat different flavor have been furnished, still in recent years, by several authors, and a couple of them are relevant to our aims.

Specifically, assume for the remainder of the paper that $\mathbb{A} = (A,+)$ is a fixed, arbitrary semigroup (unless differently specified), and let 0 be the identity of the unitization, $\mathbb{A}^{(0)}$, of $\mathbb{A}$ (cf. [Ho, p. 2]): If $\mathbb{A}$ is unital, then $\mathbb{A}^{(0)} := \mathbb{A}$; otherwise, $\mathbb{A}^{(0)}$ is the pair $(A \cup \{A\}, +)$, where we abuse notation and continue writing $+$ for the unique extension of $+$ to a binary operation on $A \cup \{A\}$ for which A serves as an identity (note that $A \notin A$, so loosely speaking we are just adjoining a distinguished element to A and extending the structure of $\mathbb{A}$ in such a way that the outcome is a monoid whose identity is the adjoined element). We denote by

$\mathfrak{p}(\mathbb{A})$ the infimum of $\text{ord}_{\mathbb{A}^{(0)}}(z)$ as z ranges over the elements of $\mathbb{A}^{(0)} \setminus \{0\}$, with the convention that $\mathfrak{p}(\mathbb{A}) := |\mathbb{N}|$ if $\mathbb{A}^{(0)} = \{0\}$, i.e. $\mathbb{A}^{(0)}$ is trivial. Then we have:

Theorem 4 (folklore)**.** *If $\mathbb{A}$ is a commutative group and X, Y are non-empty subsets of A, then $|X + Y| \ge \min(\mathfrak{p}(\mathbb{A}), |X| + |Y| - 1)$.*

Theorem 4 is another (straightforward) consequence of Kneser's theorem. While it applies to both finite and infinite *commutative* groups, an analogous result holds true for all groups:

Theorem 5 (Hamidoune-Károlyi theorem)**.** *If $\mathbb{A}$ is a group and X, Y are non-empty subsets of A, then $|X + Y| \ge \min(\mathfrak{p}(\mathbb{A}), |X| + |Y| - 1)$.*

This was first proved by Károlyi in the case of finite groups, relying on the structure theory of group extensions, by reduction to finite solvable groups in the light of the Feit-Thompson theorem, and then by Hamidoune in the general case, based on the isoperimetric method; see [K] for details.

A further result from the literature that is significant in relation to the subject matter is due to J.H.B. Kemperman [Ke], and reads as follows:

Theorem 6 (Kemperman's inequality for torsion-free groups)**.** *Let $\mathbb{A}$ be a group, and let X, Y be non-empty subsets of A. Suppose that every non-zero element of A has order $\ge |X| + |Y| - 1$. Then, $|X + Y| \ge |X| + |Y| - 1$.*

In fact, [Ke] is focused on cancellative semigroups (there simply called semigroups), and it is precisely in this framework that Kemperman establishes a series of results, related to the number of different representations of an element in a sumset, eventually leading to Theorem 6, a weak version of which will be proved in Section 5 as a corollary of our main theorem (namely, Corollary 13).

For the rest, Hamidoune and coauthors, see [CHS, Theorem 3], have proved a Cauchy-Davenport theorem for *acyclic* monoids (these are termed acyclic *semigroups* in [CHS], but they are, in fact, *monoids* in our terminology), and it would be interesting to find a common pattern between their result and those in the present work. Unluckily, the author has no clue on this for the moment (in particular, note that acyclic semigroups in [CHS] are not cancellative semigroups).

Organization.

In Section 2, we define the Cauchy-Davenport constant of a pair of sets in a semigroup and state our main results. In Section 3, we establish a few basic lemmas. Section 4 is devoted to generalized Davenport transforms and their fundamental properties. We demonstrate the central theorem of the paper (namely, Theorem 8) in Section 5, and we give a couple of applications in Section 6.

2. The statement of the main results

Keeping all the above in mind, we can now proceed to the heart of the paper.

Definition 7. For a subset Z of A, we let

$$\omega_{\mathbb{A}}(Z) := \sup_{z_0 \in Z \cap A^\times} \inf_{z \in Z \setminus \{z_0\}} \operatorname{ord}(z - z_0). \tag{1}$$

Then, given $X, Y \subseteq A$ we define $\Omega_{\mathbb{A}}(X, Y) := 0$ if either of X and Y is empty; $\Omega_{\mathbb{A}}(X, Y) := \max(|X|, |Y|)$ if $X \times Y \neq \emptyset$ and either of X and Y is infinite, and

$$\Omega_{\mathbb{A}}(X, Y) := \min(\omega_{\mathbb{A}}(X, Y), |X| + |Y| - 1)$$

otherwise, where $\omega_{\mathbb{A}}(X, Y) := \max(\omega_{\mathbb{A}}(X), \omega_{\mathbb{A}}(Y))$. We refer to $\Omega_{\mathbb{A}}(X, Y)$ as the *Cauchy-Davenport constant of* (X, Y) *relative to* $\mathbb{A}$ (again, the subscript '$\mathbb{A}$' may be omitted from the notation if there is no danger of ambiguity).

Here and later, we assume that the supremum of the empty set is 0, while its infimum is $|\mathbb{N}|$, so any pair of subsets of A has a well-defined Cauchy-Davenport constant (relative to $\mathbb{A}$). In particular, $\omega(Z)$ is zero for $Z \subseteq A$ if $Z \cap A^\times = \emptyset$. However, this is not the case, for instance, when $Z \neq \emptyset$ and $\mathbb{A}$ is a group, which is the "moral basis" for the following non-trivial bound:

Theorem 8. *Suppose* $\mathbb{A}$ *is cancellative and let* X, Y *be subsets of* A *such that* $\langle Y \rangle$ *is commutative. Then,* $|X + Y| \geq \min(\omega(Y), |X| + |Y| - 1)$ *if both of* X *and* Y *are finite and non-empty, and* $|X + Y| \geq \Omega(X, Y)$ *otherwise.*

Theorem 8 represents the central contribution of the paper. Not only it extends the Cauchy-Davenport theorem to the broader and more abstract setting of semigroups, but it also provides a strengthening and a generalization of Theorem 4 in view of the following:

Lemma 9. *If* Z *is a subset of* A *such that* $Z \cap A^\times \neq \emptyset$*, then* $\omega(Z) \geq \mathfrak{p}(\mathbb{A})$.

Proof. Having $Z \cap A^\times \neq \emptyset$, pick $z_0 \in Z \cap A^\times$. If Z is a singleton, the assertion is trivial since then $\inf_{z \in Z \setminus \{z_0\}} \operatorname{ord}(z - z_0) = |\mathbb{N}|$. Otherwise, taking $z \in Z \setminus \{z_0\}$ gives $\operatorname{ord}(z - z_0) \geq \mathfrak{p}(\mathbb{A})$ by the definition of $\mathfrak{p}(\mathbb{A})$. □

Lemma 9 applies, on the level of groups, to *any* non-empty subset (see Corollary 12 below), and the stated inequality is strict in significant cases: For a concrete example, pick $k, q \in \mathbb{N}^+$ and set $m := qk$ and $X := \{(1 + ik) \bmod m : i = 1, \ldots, q\}$. Then observe that $|2X| = \Omega_{\mathbb{Z}/m\mathbb{Z}}(X, X) = q$, while $\mathfrak{p}(\mathbb{Z}/m\mathbb{Z})$ is the smallest prime, say p, dividing m, in such a way that $\mathfrak{p}(\mathbb{Z}/m\mathbb{Z})$ is "much" smaller than $\Omega_{\mathbb{Z}/m\mathbb{Z}}(X, X)$ if p is "much" smaller than q.

Theorem 8 can be "symmetrized" and further strengthened in the case where each summand generates a commutative subsemigroup, leading to the following corollaries, whose proofs are straightforward in the light of Definition 7:

Corollary 10. *Assume $\mathbb{A}$ is cancellative and let X, Y be subsets of A such that $\langle X \rangle$ is commutative. Then, $|X + Y| \ge \min(\omega(X), |X| + |Y| - 1)$ if both of X and Y are finite and non-empty, and $|X + Y| \ge \Omega(X, Y)$ otherwise.*

Corollary 11. *If $\mathbb{A}$ is cancellative and X, Y are subsets of A such that both of $\langle X \rangle$ and $\langle Y \rangle$ are commutative, then $|X + Y| \ge \Omega(X, Y)$.*

Moreover, the result specializes to groups as follows:

Corollary 12. *If $\mathbb{A}$ is a group and X, Y are non-empty subsets of A such that $\langle Y \rangle$ is commutative. Then, $|X + Y| \ge \min(\omega(Y), |X| + |Y| - 1)$, where*

$$\omega(Y) = \sup_{y_0 \in Y} \inf_{y \in Y \setminus \{y_0\}} \operatorname{ord}(y - y_0),$$

and in particular $\omega(Y) = \max_{y_0 \in Y} \inf_{y \in Y \setminus \{y_0\}} \operatorname{ord}(y - y_0)$ if Y is finite.

Proof. Immediate by Theorem 8. For, on the one hand $\mathbb{A}$ being a group implies $Y = Y \cap A^\times$, and on the other, a supremum over a finite set is a maximum. □

The next corollary is now a *partial* generalization of Theorem 6 to cancellative semigroups: its proof is straightforward by Corollary 11 and Lemma 9. Here, we say that $\mathbb{A}$ is torsion-free if $\mathfrak{p}(\mathbb{A})$ is infinite (the analogous notion in group theory is, in fact, a special case).

Corollary 13. *If $\mathbb{A}$ is cancellative, and if X, Y are non-empty subsets of A such that every element of $A \setminus \{0\}$ has order $\ge |X| + |Y| - 1$ (this is especially the case when $\mathbb{A}$ is torsion-free) and either of $\langle X \rangle$ and $\langle Y \rangle$ is abelian, then $|X + Y| \ge |X| + |Y| - 1$.*

Theorem 8 is proved in Section 5. The argument is inspired by the transformation proof originally used for Theorem 1 by Davenport in [D1]. This leads to the definition of what we call a *generalized Davenport transform*. The author is not aware of an earlier use of the same technique in the literature, all the more in relation to semigroups. With few exceptions, remarkably including [HR] and A.G. Vosper's original proof of his famous theorem on critical pairs [V], even the "classical" Davenport transform has not been greatly considered by practitioners in the area, especially in comparison with similar "technology" such as the Dyson transform [N, p. 42].

Remark 14. A couple of things are worth mentioning before proceeding. While every *commutative* cancellative semigroup embeds as a subsemigroup into a group (as it follows from the standard construction of the group of fractions of a commutative monoid; see [B1, Chapter I, Section 2.4]), nothing similar is true in the *non*-commutative case.

This has to do with a well-known question in semigroup theory, first answered by A.I. Mal'cev in [M], and serves as a fundamental motivation for the present paper, in that it shows that the study of sumsets in cancellative semigroups cannot be systematically reduced, in the absence of commutativity, to the case of groups (at least, not in any obvious way). Incidentally, note that the semigroup in Mal'cev's example is finitely generated.

On another hand, it is true that every cancellative semigroup can be embedded into a cancellative monoid (through the unitization process mentioned in the comments preceding the statement of Theorem 4, in Section 1), so that, for the *specific purposes of the manuscript*, we could have assumed in most of our statements that the "ambient" is a monoid (rather than a semigroup), but we did differently because the assumption is not really necessary. We will see, however, that certain parts take a simpler and more natural form when an identity is made available somehow, this being especially the case with a few lemmas in Section 3 and the proof of Theorem 8.

We provide two applications of Theorem 8 (others will be investigated in future work): The first is a generalization of Theorem 2, the second is an improvement of a previous result by Ø.J. Rødseth [R, Section 6] based on Hall's "marriage theorem". As for the former (which is stated below), we will use the following specific notation: Given $m \in \mathbb{N}^+$ and a non-empty $Z \subseteq \mathbb{Z}/m\mathbb{Z}$, we let

$$\delta_Z := \min_{z_0 \in Z} \max_{z \in Z \setminus \{z_0\}} \gcd(m, z - z_0) \tag{2}$$

if $|Z| \ge 2$, and $\delta_Z := 1$ otherwise. Then we have:

Corollary 15. *For an integer $m \ge 1$ let X and Y be non-empty subsets of $\mathbb{Z}/m\mathbb{Z}$ and define $\delta := \min(\delta_X, \delta_Y)$. Then,*

$$|X + Y| \ge \min(\delta^{-1} m, |X| + |Y| - 1).$$

In particular, $|X + Y| \ge \min(m, |X| + |Y| - 1)$ if there exists $y_0 \in Y$ such that m is prime with $y - y_0$ for each $y \in Y \setminus \{y_0\}$ (or dually with X in place of Y).

Corollary 15 contains Chowla's theorem (Theorem 2) as a special case: With the same notation as above, it is enough to assume that the identity of $\mathbb{Z}/m\mathbb{Z}$ belongs to Y, and $\gcd(m, y) = 1$ for each non-zero $y \in Y$. Moreover, it is clear from (2) that the corollary is a strengthening of Pillai's theorem (Theorem 3).

Many questions arise. Most notably: Is it possible to further extend Corollary 11 in such a way to get rid of the assumption that any of the summands generates a commutative subsemigroup? This sounds particularly significant, for a positive answer would provide a comprehensive generalization of about all the extensions of the Cauchy-Davenport theorem reviewed in Section 1, and remarkably of Theorems 5 and 6.

3. Preliminaries

This section collects basic results used later to introduce the generalized Davenport transforms and prove Theorem 8. Some proofs are direct and standard (and thus omitted without further explanation), but we have no reference to anything similar in the context of semigroups, so we include the statements here for completeness.

Lemma 16. *Pick $n \in \mathbb{N}^+$ and subsets $X_1, Y_1, \ldots, X_n, Y_n$ of A such that $X_i \subseteq Y_i$ for each i. Then, $\sum_{i=1}^n X_i \subseteq \sum_{i=1}^n Y_i$ and $\left|\sum_{i=1}^n X_i\right| \le \left|\sum_{i=1}^n Y_i\right|$.*

Lemma 17. *Assume $\mathbb{A}$ is cancellative and pick an integer $n \ge 2$ and non-empty $X_1, \ldots, X_n \subseteq A$. Then, $\left|\sum_{i=2}^n X_i\right| \le \left|\sum_{i=1}^n X_i\right|$ and $\left|\sum_{i=1}^{n-1} X_i\right| \le \left|\sum_{i=1}^n X_i\right|$.*

For the next lemma, whose proof is straightforward by a routine induction, we assume that $0 \cdot \kappa := \kappa \cdot 0 := 0$ for every cardinal κ.

Lemma 18. *For $n \in \mathbb{N}^+$ and $X_1, \ldots, X_n \subseteq A$ it holds $\left|\sum_{i=1}^n X_i\right| \le \prod_{i=1}^n |X_i|$.*

Let $X, Y \subseteq A$. No matter if $\mathbb{A}$ is cancellative, nothing similar to Lemmas 17 and 18 applies, in general, to the difference set $X - Y$, in the sense that this can be infinite even if both of X and Y are finite. On another hand, we get by symmetry and Lemma 17 that, in the presence of cancellativity, the cardinality of the sumset $X + Y$ is preserved under translation, namely $|z + X + Y| = |X + Y + z| = |X + Y|$ for every $z \in A$. This is a point in common with the case of groups, save for the fact that we cannot profit from it, at least in general, to "normalize" either of X and Y in such a way as to contain some distinguished element of A.

Lemma 19. *Let X and Y be subsets of A. The following are equivalent:*

(i) $X + 2Y \subseteq X + Y$.

(ii) $X + nY \subseteq X + Y$ *for all* $n \in \mathbb{N}^+$.

(iii) $X + \langle Y \rangle = X + Y$.

Proof. Points (ii) and (iii) are clearly equivalent, as $X+\langle Y\rangle = \bigcup_{n=1}^{\infty}(X+nY)$, and (i) is obviously implied by (ii). Thus, we are left to prove that (ii) follows from (i), which is immediate (by induction) using that, if $X+nY \subseteq X+Y$ for some $n \in \mathbb{N}^+$, then $X+(n+1)Y = (X+nY)+Y \subseteq (X+Y)+Y = X+2Y \subseteq X+Y$. □

The above result is as elementary as central in the plan of the paper, in that the properties of the generalized Davenport transform used later, in Section 5, for the proof of Theorem 8 strongly depend on it.

On another hand, the following lemma shows that, in reference to Theorem 8, there is no loss of generality in assuming that the ambient semigroup is unital, for the fact that any semigroup embeds as a subsemigroup into its unitization.

Lemma 20. *Let $(B,\star)$ be a semigroup, φ an injective function from A to B such that $\varphi(z_1+z_2) = \varphi(z_1)\star\varphi(z_2)$ for all $z_1, z_2 \in A$, and $X_1, \ldots, X_n \subseteq A$ ($n \in \mathbb{N}^+$). Then, $|X_1+\cdots+X_n| = |\varphi(X_1)\star\cdots\star\varphi(X_n)|$.*

We close the section with a few properties of units. Here and later, if X is a subset of A, we use $\mathrm{C}_{\mathbb{A}}(X)$ for the set of all $z \in A$ such that $z+x = x+z$ for every $x \in X$ (this is called the centralizer of X in $\mathbb{A}$).

Lemma 21. *Let $\mathbb{A}$ be a monoid, X a subset of A, and z a unit of $\mathbb{A}$ with inverse $\tilde{z}$. Then, the following conditions hold:*

(i) $X-z = X+\tilde{z}$, $-z+X = \tilde{z}+X$ *and* $|-z+X| = |X-z| = |X|$.

(ii) *If $z \in \mathrm{C}_{\mathbb{A}}(X)$, then $\tilde{z} \in \mathrm{C}_{\mathbb{A}}(X)$; in addition to this, $\langle X-z\rangle$ and $\langle -z+X\rangle$ are commutative if $\langle X\rangle$ is commutative.*

Proof. (i) By symmetry, it suffices to prove that $X-z = X+\tilde{z}$ and $|X-z| = |X|$. As for the first identity, notice that $w \in X-z$ if and only if there exists $x \in X$ such that $w+z = x$, which is equivalent to $x+\tilde{z} = (w+z)+\tilde{z} = w$, namely $w \in X+\tilde{z}$. In order to conclude, it is then sufficient to observe that the function $A \to A : \xi \mapsto \xi+\tilde{z}$ is a bijection.

(ii) Pick $z \in \mathrm{C}_{\mathbb{A}}(X)$ and $x \in X$. It is then seen that $x+\tilde{z} = \tilde{z}+x$ if and only if $x = (x+\tilde{z})+z = \tilde{z}+x+z$, and this is certainly true as our standing assumptions imply $\tilde{z}+x+z = \tilde{z}+z+x = x$. It follows that $\tilde{z} \in \mathrm{C}_{\mathbb{A}}(X)$.

Suppose now that $\langle X\rangle$ is a commutative semigroup and let $v, w \in \langle X-z\rangle$. By point (i) above, there exist $k, \ell \in \mathbb{N}^+$ and $x_1, \ldots, x_k, y_1, \ldots, y_\ell \in X$ such that $v = \sum_{i=1}^{k}(x_i+\tilde{z})$ and $w = \sum_{i=1}^{\ell}(y_i+\tilde{z})$, with the result that $v+w = w+v$ by induction on $k+\ell$ and the observation that for all $u_1, u_2 \in X$ it holds

$$(u_1+\tilde{z})+(u_2+\tilde{z}) = u_1+u_2+2\tilde{z} = u_2+u_1+2\tilde{z} = (u_2+\tilde{z})+(u_1+\tilde{z}),$$

where we use that $\tilde{z} \in \mathrm{C}_{\mathbb{A}}(X)$, as proved before, and $\langle X\rangle$ is commutative. Hence, $\langle X-z\rangle$ is commutative too, which completes the proof by symmetry. □

Remark 22. There is a subtleness in Definition 7 which we have so far intentionally overlooked, but should be remarked. For, suppose that $\mathbb{A}$ is a monoid and pick $x, y \in A$. In principle, $x - y$ and $-y + x$ are not elements of A: In fact, they are (difference) sets, and no other meaningful interpretation is possible a priori. However, if y is a unit of $\mathbb{A}$ and $\tilde{y}$ is the inverse of y, then $x - y = \{x + \tilde{y}\}$ and $-y + x = \{\tilde{y} + x\}$ by point (i) of Lemma 21, and we are allowed to identify $x - y$ with $x + \tilde{y}$ and $-y + x$ with $\tilde{y} + x$, which is useful in many situations.

4. The Davenport transform revisited

As mentioned in Section 2, Davenport's proof [D1, Statement A] of Theorem 1 is a transformation proof. For $\mathbb{A}$ a *commutative group*, the idea is to map a pair (X, Y) of non-empty subsets of A to a new pair (X, Y'), which is smaller than (X, Y) in an appropriate sense, and specifically such that

$$|Y'| < |Y|, \quad |X + Y'| + |Y| \le |X + Y| + |Y'|.$$

We then refer to (X, Y') as a Davenport transform of (X, Y); see, for instance, [HR]. For this to be possible, the classical approach requires that $X + 2Y \not\subseteq X + Y$ and $0 \in Y$, so that $|Y| \ge 2$.

As expected, many difficulties arise when attempting to adapt the same approach to semigroups, all the more if these are non-commutative. Even the possibility of embedding a semigroup into a monoid does not resolve anything, since the fundamental problem is that, contrary to the case of groups, cardinality is not preserved "under subtraction". Namely, if $\mathbb{A}$ is an arbitrary monoid with identity 0 (as intended for the remainder of the section, unless differently stated), X is a subset of A, and $z \in A$, then $|X|$, $|X - z|$ and $|-z + X|$ can be greatly different from each other, even supposing that $\mathbb{A}$ is cancellative; cf. point (i) of Lemma 21. Thus, unless $\mathbb{A}$ is a group or, more generally, embeds as a submonoid into a group, we are not allowed to assume, for instance, that 0 is in Y by picking an arbitrary element $y_0 \in Y$ and replacing (X, Y) with the "shifted" pair $(X + y_0, -y_0 + Y)$; cf. the comments following Lemma 18.

In fact, the primary goal of this section is to show that, in spite of these issues, Davenport's original ideas can be extended and used for a proof of Theorem 8.

To start with, let X and Y be subsets of A such that $mX + 2Y \not\subseteq X + Y$ for some positive integer m. For the sake of brevity, define

$$Z := (mX + 2Y) \setminus (X + Y).$$

Our assumptions give $Z \ne \emptyset$. So fix $z \in Z$, and take $x_z \in (m-1)X$ and $y_z \in Y$ for which $z \in x_z + X + Y + y_z$, where $0X := \{0\}$. Finally, set

$$\tilde{Y}_z := \{y \in Y : z \in x_z + X + Y + y\}, \quad Y_z := Y \setminus \tilde{Y}_z. \tag{3}$$

We refer to (X, Y_z) as a *generalized Davenport transform* of (X, Y) (relative to z), and based on this notation we have:

Proposition 23. *If $Y_z \ne \emptyset$, then the triple $(X, Y_z, \tilde{Y}_z)$ satisfies the following:*

(i) *Y_z and $\tilde{Y}_z$ are non-empty disjoint proper subsets of Y, and $\tilde{Y}_z = Y \setminus Y_z$.*

(ii) *If $\mathbb{A}$ is cancellative, then $(x_z + X + Y_z) \cup (z - \tilde{Y}_z) \subseteq x_z + X + Y$.*

(iii) *If $\langle Y \rangle$ is commutative, then $(x_z + X + Y_z) \cap (z - \tilde{Y}_z) = \emptyset$.*

(iv) *If $\mathbb{A}$ is cancellative, then $|z - \tilde{Y}_z| \ge |\tilde{Y}_z|$.*

(v) *$|X+Y| + |Y_z| \ge |X + Y_z| + |Y|$ if $\mathbb{A}$ is cancellative and $\langle Y \rangle$ is commutative.*

Proof. (i) Y_z is non-empty by hypothesis, while $\tilde{Y}_z$ is non-empty since $y_z \in \tilde{Y}_z$ by construction. In addition to this, (3) gives $Y_z, \tilde{Y}_z \subseteq Y$ and $Y_z \cap \tilde{Y}_z = \emptyset$, so that $Y \setminus Y_z = Y \setminus (Y \setminus \tilde{Y}_z) = \tilde{Y}_z$ and $Y_z, \tilde{Y}_z \subsetneq Y$.

(ii) Since $Y_z \subseteq Y$ by point (i) above, we have $x_z + X + Y_z \subseteq x_z + X + Y$ by Lemma 16. On the other hand, if $w \in z - \tilde{Y}_z$, then there exists $y \in \tilde{Y}_z$ such that $z = w + y$. But $y \in \tilde{Y}_z$ implies by (3) that $z = \tilde{w} + y$ for some $\tilde{w} \in x_z + X + Y$, hence $w = \tilde{w}$ by cancellativity, namely $w \in x_z + X + Y$.

(iii) Assume the contrary and let $w \in (x_z + X + Y_z) \cap (z - \tilde{Y}_z)$. There then exist $x \in X$, $y_1 \in Y_z$ and $y_2 \in \tilde{Y}_z$ such that $w = x_z + x + y_1$ and $z = w + y_2$. As $\langle Y \rangle$ is commutative, it follows that $z = x_z + x + y_1 + y_2 = x_z + x + y_2 + y_1$, which in turn implies $y_1 \in \tilde{Y}_z$ by (3), since $Y_z, \tilde{Y}_z \subseteq Y$ by point (i). This is, however, absurd, because $Y_z \cap \tilde{Y}_z = \emptyset$ by the same point (i).

(iv) We have from (3) that for each $y \in \tilde{Y}_z$ there exists $w \in x_z + X + Y$ such that $z = w + y$, hence $w \in z - \tilde{Y}_z$. On another hand, since $\mathbb{A}$ is cancellative, we cannot have $w + y_1 = w + y_2$ for some $w \in A$ and distinct $y_1, y_2 \in \tilde{Y}_z$. Thus, $\tilde{Y}_z$ embeds as a set into $z - \tilde{Y}_z$, with the result that $|z - \tilde{Y}_z| \ge |\tilde{Y}_z|$.

(v) As $\mathbb{A}$ is cancellative and X is non-empty (otherwise $Z = \emptyset$), we have by symmetry and Lemma 17 that $|X + Y| \ge \max(|X|, |Y|)$. This implies the claim if Y is infinite, since then either $|X+Y| > |Y|$, and hence

$$|X+Y| + |Y_z| = |X| = |X + Y_z| + |Y|,$$

or $|X+Y| = |Y|$, and accordingly

$$|X + Y_z| + |Y_z| = |Y| = |X + Y_z| + |Y|.$$

Notice here that we are using the axiom of choice (assumed in the background as part of our foundations) to say that $|X+Y| = \max(|X|,|Y|)$ if both of X and Y are infinite. So we are left with the case when Y is finite, for which the inclusion-exclusion principle, points (ii)-(iv) and Lemmas 17 and 18 give, by symmetry, that

$$|X+Y| = |x_z + X + Y| \geq |x_z + X + Y_z| + |z - \tilde{Y}_z| =$$
$$= |X + Y_z| + |z - \tilde{Y}_z| \geq |X + Y_z| + |\tilde{Y}_z|.$$

But $\tilde{Y}_z = Y \setminus Y_z$ and $Y_z \subseteq Y$ by point (i) above, so at the end we have $|X+Y| \geq |X+Y_z| + |Y| - |Y_z|$, and the proof is thus complete. □

Remark 24. To apply the generalized Davenport transform to Theorem 8, it will be enough to consider the case where $m = 1$, for which it is easily seen that $0 \in Y_z$ if $0 \in Y$ (we continue with the notation from above), as otherwise we would have $z \in X+Y$, in contradiction to the fact that $z \in (X+2Y)\setminus(X+Y)$. However, it seems intriguing that the same machinery can be used, at least in principle, even if $m \geq 2$ insofar as there is a way to prove that Y_z is non-empty.

5. The proof of the main theorem

Proposition 23 is used here to establish the main contribution of the paper.

Proof of Theorem 8. Since every semigroup embeds as a subsemigroup into its unitization, and the unitization of a cancellative semigroup is cancellative in its own right, Lemma 20 and Definition 7 imply that there is no loss of generality in assuming, as we do, that $\mathbb{A}$ is unital.

Thus, we let 0 denote the identity of $\mathbb{A}$, and we suppose by contradiction that the theorem is false. There then exists a pair (X,Y) of subsets of A for which $\langle Y \rangle$ is abelian and $|X+Y| < \min(\omega(Y), |X|+|Y|-1)$. Then,

$$2 \leq |X|, |Y| < |\mathbb{N}|. \tag{4}$$

In fact, if either of X and Y is a singleton or infinite then $|X+Y| = \max(|X|,|Y|)$, and Definition 7 gives $|X+Y| = \Omega(X,Y)$, contradicting the standing assumptions. It follows from (1) and (4) that

$$|X+Y| < \sup_{y_0 \in Y \cap A^\times} \inf_{y \in Y \setminus \{y_0\}} \operatorname{ord}(y - y_0), \quad |X+Y| \leq |X|+|Y|-2. \tag{5}$$

Again without loss of generality, we also take $|X|+|Y|$ to be minimal over the pairs of subsets of A for which, in particular, (4) and (5) are presumed to be true.

Now, since $|X+Y|$ is finite by (4) and Lemma 18, we get by (5) and the same equation (4) that there exists $\tilde{y}_0 \in Y \cap A^\times$ such that

$$|X+Y| < \inf_{y \in Y \setminus \{\tilde{y}_0\}} \operatorname{ord}(y - \tilde{y}_0) = \min_{y \in Y \setminus \{\tilde{y}_0\}} \operatorname{ord}(y - \tilde{y}_0). \tag{6}$$

So letting $W_0 := Y - \tilde{y}_0$ implies by (5) and (6) that

$$|X+W_0| < \min_{w \in W_0 \setminus \{0\}} \operatorname{ord}(w), \quad |X+W_0| \le |X| + |W_0| - 2. \tag{7}$$

In fact, on the one hand $|Y - \tilde{y}_0| = |Y|$ and $|X+Y-\tilde{y}_0| = |X+Y|$ by point (i) of Lemma 21, and on the other, $y \in Y \setminus \{\tilde{y}_0\}$ only if $y - \tilde{y}_0 \in (Y - \tilde{y}_0) \setminus \{0\}$, while $w \in (Y - \tilde{y}_0) \setminus \{0\}$ only if $w + \tilde{y}_0 \in Y \setminus \{\tilde{y}_0\}$ (recall Remark 22). We claim that

$$Z := (X + 2W_0) \setminus (X + W_0) \ne \emptyset. \tag{8}$$

For, suppose the contrary. Then, $X + W_0 = X + \langle W_0 \rangle$ by Lemma 19, so that

$$|X+W_0| = |X + \langle W_0 \rangle| \ge |\langle W_0 \rangle| \ge \max_{w \in W_0} \operatorname{ord}(w) \ge \min_{w \in W_0 \setminus \{0\}} \operatorname{ord}(w),$$

where we use, in particular, Lemma 17 for the first inequality and the fact that $|W_0| \ge 2$ for the last one. However, this contradicts (7), so (8) is proved.

Pick $z \in Z$ and let (X, W_0') be a generalized Davenport transform of (X, W_0) relative to z. As $\langle Y \rangle$ is a commutative subsemigroup of $\mathbb{A}$ (by hypothesis), the same is true for $\langle W_0 \rangle$, by point (ii) of Lemma 21. Further, 0 is in W_0, and thus

$$0 \in W_0' \ne \emptyset, \quad W_0' \subsetneq W_0, \tag{9}$$

when taking into account Remark 24 and point (i) of Proposition 23. Therefore, point (v) of the same Proposition 23 yields, together with (7), that

$$|X + W_0'| + |W_0| \le |X + W_0| + |W_0'| \le |X| + |W_0| - 2 + |W_0'|,$$

which means, since $|W_0| = |Y - \tilde{y}_0| = |Y| < |\mathbb{N}|$ by (4) and the above, that

$$|X + W_0'| \le |X| + |W_0'| - 2. \tag{10}$$

It follows from (9) that $1 \le |W_0'| < |W_0|$, and in fact $|W_0'| \ge 2$, as otherwise we would have $|X| = |X + W_0'| \le |X| - 1$ by (10), in contradiction to the fact that $|X| < |\mathbb{N}|$ by (4). To summarize, we have found that

$$2 \le |W_0'| < |W_0| < |\mathbb{N}|. \tag{11}$$

Furthermore, (7) and (9) entail that

$$|X + W_0'| \le |X + W_0| < \min_{w \in W_0' \setminus \{0\}} \operatorname{ord}(w), \tag{12}$$

where we use that $\min(C_1) \le \min(C_2)$ if C_1 and C_2 are sets of cardinal numbers with $C_2 \subseteq C_1$. Thus, since $0 \in W_0' \cap A^\times$, we get by (12) that

$$|X + W_0'| < \sup_{w_0 \in W_0' \cap A^\times} \min_{w \in W_0' \setminus \{w_0\}} \operatorname{ord}(w), \tag{13}$$

which however contradicts, by (4), (10) and (11), the minimality of $|X| + |Y|$, for the fact that $|W_0'| < |W_0| = |Y|$, and hence $|X| + |W_0'| < |X| + |Y|$. □

6. A couple of applications

First, we show how to use Theorem 8 to prove the extension of Chowla's theorem for composite moduli mentioned in Section 2.

Proof of Corollary 15. The claim is trivial if either of X and Y is a singleton. Otherwise, since $\mathbb{Z}/m\mathbb{Z}$ is a commutative finite group and $\operatorname{ord}(z - z_0) = m/\gcd(m, z - z_0)$ for all $z, z_0 \in \mathbb{Z}/m\mathbb{Z}$, we get by Corollary 12 that $|X + Y| \ge \min(\omega(Y), |X| + |Y| - 1)$, where

$$\omega(Y) = \max_{y_0 \in Y} \min_{y \in Y \setminus \{y_0\}} \operatorname{ord}(y - y_0) = \max_{y_0 \in Y} \min_{y \in Y \setminus \{y_0\}} \frac{m}{\gcd(m, y - y_0)} = \delta_Y^{-1} m.$$

In an entirely similar way, it is found, in view of Corollary 10, that

$$|X + Y| \ge \min(\delta_X^{-1} m, |X| + |Y| - 1).$$

This concludes the proof, considering that $\delta_Y = 1$ if there exists $y_0 \in Y$ such that $\gcd(m, y - y_0) = 1$ for every $y \in Y \setminus \{y_0\}$ (and dually with X). □

We now use P. Hall's theorem on distinct representatives [H] to say something on the "localization" of elements in a sumset.

Theorem 25 (Hall's theorem). *Let $S_1, \ldots, S_n$ be arbitrary sets, $n \in \mathbb{N}^+$. There then exist (pairwise) distinct elements $s_1 \in S_1, \ldots, s_n \in S_n$ if and only if for each $k = 1, \ldots, n$ the union of any k of $S_1, \ldots, S_n$ contains at least k elements.*

More precisely, suppose $\mathbb{A}$ is a cancellative semigroup and let X, Y be nonempty finite subsets of A such that $|X + Y| < \omega(Y)$. Clearly, this implies $Y \cap A^\times \neq \emptyset$. Define $k := |X|$ and $\ell := |Y|$, and let $x_1, \ldots, x_k$ be a numbering of X and $y_1, \ldots, y_\ell$ a numbering of Y. Then consider the k-by-ℓ matrix, say $\alpha(X, Y)$, whose entry in the i-th row and j-th column is $x_i + y_j$. Any element of $X + Y$ appears in $\alpha(X, Y)$, and viceversa any entry of $\alpha(X, Y)$ is an element of $X + Y$. Also, Theorem 8 and our hypotheses give $|X + Y| \ge k + \ell - 1$. So it is natural to try to get some information about where in the matrix $\alpha(X, Y)$ it is possible to find $k + \ell - 1$ distinct elements of $X + Y$.

In this respect we have the following proposition, whose proof is quite similar to the one of a weaker result in [R, Section 6], which is, in turn, restricted to the case of a group of prime order:

Proposition 26. *Assume that $\langle Y\rangle$ is commutative and let Z be any subset of $X+Y$ of size $\ell-1$, for instance $Z = x_1 + \{y_1, \ldots, y_{\ell-1}\}$. Then, we can choose one element from each row of $\alpha(X,Y)$ in such a way that Z and these elements form a subset of $X+Y$ of size $k+\ell-1$.*

Proof. For $i = 1, \ldots, k$ let $Z_i := (x_i + Y) \setminus Z$, and note that Z_i is a subset of the i-th row of $\alpha(X,Y)$. So $Z_{i_1} \cup \cdots \cup Z_{i_h} = (\{x_{i_1}, \ldots, x_{i_h}\} + Y) \setminus Z$ for any positive integer $h \le k$ and all distinct indexes $i_1, \ldots, i_h \in \{1, \ldots, k\}$. It follows that

$$|Z_{i_1} \cup \cdots \cup Z_{i_h}| \ge |\{x_{i_1}, \ldots, x_{i_h}\} + Y| - |Z| \ge h + \ell - 1 - (\ell - 1) = h,$$

where we combine Theorem 8 with the fact that

$$|\{x_{i_1}, \ldots, x_{i_h}\} + Y| \le |X + Y| < \omega(Y),$$

as is implied by Lemma 16 and the assumption that $|X+Y| < \omega(Y)$. Then, as a consequence of Hall's theorem, we can find k distinct elements $z_1 \in Z_1, \ldots, z_k \in Z_k$, and these, together with the $\ell-1$ elements of Z, provide a total of $k+\ell-1$ distinct elements of $X+Y$, as $Z \cap Z_1 = \cdots = Z \cap Z_k = \emptyset$ (by construction). □

Acknowledgements

The author is grateful to Andrea Gagna (Università di Milano), Alain Plagne (Centre de mathématiques Laurent Schwartz, École polytechnique), Carlo Sanna (Università di Torino), and an anonymous referee for comments that helped to improve this paper significantly.

References

[B1] Bourbaki, N.: *Algèbre, Chapitres 1 à 3*, Éléments de mathématique II, Springer-Verlag, Berlin, 2006 (2nd revised ed.).

[B2] Bourbaki, N.: *Théorie des ensembles*, Éléments de mathématique I, Springer-Verlag, Berlin, 2006 (reprint ed.).

[C] Cauchy, A.-L.: *Recherches sur les nombres*, J. École Polytech. **9** (1813), 99–116 (reproduced in *Oeuvres*, Série 2, Tome 1, 39–63).

[CHS] Cilleruelo, A.L. – Hamidoune, Y.O. – Serra, O.: *Addition theorems in acyclic semigroups*, 99–104 in 'Additive number theory', Springer, 2010.

[Ch] Chowla, I.: *A theorem on the addition of residue classes: Application to the number $\Gamma(k)$ in Waring's problems*, Proc. Indian Acad. Sc. (A), **2** (1935), 242–243.

[D1] Davenport, H.: *On the addition of residue classes*, J. Lond. Math. Soc. **10** (1935), 30–32.

[D2] Davenport, H.: *A historical note*, J. Lond. Math. Soc. **22** (1947), 100–101.

[G] Geroldinger, A.: *Additive Group Theory and Non-unique Factorizations*, 1–86 in 'Combinatorial Number Theory and Additive Group Theory', Springer, 2009.

[H] Hall, P.: *On representatives of subsets*, J. Lond. Math. Soc. **10** (1935), 26–30.

[HR] Hamidoune, Y.O. – Rødseth, Ø.J.: *An inverse theorem mod p*, Acta Arith. **92** (2000), 251–262.

[Ho] Howie, J.M.: *Fundamentals of semigroup theory*, Clarendon Press, 1995.

[K] Károlyi, G.: *The Cauchy-Davenport theorem in group extensions*, L'Enseignement Mathématique **51** (2005), 239–254.

[Ke] Kemperman, J.H.B.: *On complexes in a semigroup*, Indag. Math. **18** (1956), 247–254.

[M] Mal'cev, A.I.: *On the immersion of an algebraic ring into a field*, Math. Annalen (1)**113** (1937), 686–691.

[N] Nathanson, M.B.: *Additive Number Theory. Inverse Problems and the Geometry of Sumsets*, GTM **165**, Springer, 1996.

[P] Pillai, S.S.: *Generalization of a theorem of Davenport on the addition of residue classes*, Proc. Indian Acad. Sc. (A), (3)**6** (1937), 179–180.

[R] Rødseth, Ø.J.: *Sumsets mod p*, Skr. K. Nor. Vidensk. Selsk. (Trans. R. Norw. Soc. Sci. Lett.), **4** (2006), 1–10.

[Ru] Ruzsa, I.Z.: *Sumsets and structure*, 87–210 in 'Combinatorial Number Theory and Additive Group Theory', Springer, 2009.

[V] Vosper, A.G.: *The critical pairs of subsets of a group of prime order*, J. Lond. Math. Soc. **31** (1956), 200–205.

Received May 13, 2013
Accepted July 29, 2013

Salvatore Tringali
Laboratoire Jacques-Louis Lions (LJLL)
Université Pierre et Marie Curie (UPMC) - Paris 6
4 place Jussieu, 75005 Paris cedex 05, France.
E-mail: tringali@ann.jussieu.fr