NOTES ON SEPARABILITY

1. General observations

Let L and K be languages. We write $L|K$ if there is a regular R with $L \subseteq R$ and $R \cap K = \emptyset$. For a rational transduction $T \subseteq \Sigma^* \times \Gamma^*$, we define

$$TL = \{v \in \Gamma^* \mid \exists (u,v) \in T,\ u \in L\}, \quad T^{-1} = \{(v,u) \in \Gamma^* \times \Sigma^* \mid (u,v) \in T\}.$$

Lemma 1. *Let T be a rational transduction. Then $L|TK$ if and only if $T^{-1}L|K$.*

Proof. Suppose $L \subseteq R$ and $R \cap TK = \emptyset$ for some regular R. Then clearly $T^{-1}L \subseteq T^{-1}R$ and $T^{-1}R \cap K = \emptyset$. Therefore, the regular set $T^{-1}R$ witnesses $T^{-1}L|K$. Conversely, if $T^{-1}L|K$, then $K|T^{-1}L$ and hence, by the first direction, $(T^{-1})^{-1}K|L$. Since $(T^{-1})^{-1} = T$, this reads $TK|L$ and thus $L|TK$. $\square$

Proposition 2. *Let $\mathcal{C}$ be a full trio generated by the language G (i.e. it consists of languages TG for rational transductions T). Then regular separability for $\mathcal{C}$ can be reduced to the following problem:*

Given: *A language L from $\mathcal{C}$.*
Question: *Does $L|G$?*

Proof. Since $\mathcal{C}$ is generated by G, the input for regular separability for $\mathcal{C}$ comprises two rational transductions T_1 and T_2 and we are asked whether $T_1G|T_2G$. According to lemma 1, the latter is equivalent to $T_2^{-1}T_1G|G$. Since $T_2^{-1}T_1$ is a rational transduction, $T_2^{-1}T_1G$ is a member of $\mathcal{C}$ and this is an instance as required. $\square$

2. VASS languages and CCF/BPP languages

Let $D \subseteq \{a, \bar{a}\}^*$ be the one-sided Dyck language over one pair of parentheses. If $f_i \colon \{a, \bar{a}\}^* \to \{a_i, \bar{a}_i\}^*$ is the morphism with $f_i(a) = a_i$ and $f_i(\bar{a}) = \bar{a}_i$, then we define $D_n = f_1(D) \sqcup\!\sqcup \cdots \sqcup\!\sqcup f_n(D)$, where $\sqcup\!\sqcup$ is the shuffle operator. Observe that D_n is a CCF language (i.e. BPP language). Now proposition 2 tells us that regular separability of VASS languages can be reduced to the following problem:

Given: VASS language L and $n \in \mathbb{N}$.
Question: Does $L|D_n$?

Recall that every VASS language can be written as $h(g^{-1}(D_n) \cap R)$ for some regular R and alphabetic homomorphisms g, h. Now lemma 1 tells us that $h(g^{-1}(D_n) \cap R)|D_n$ iff $g^{-1}(D_n) \cap R|h^{-1}(D_n)$. Since $g^{-1}(D_n)$ and $h^{-1}(D_n)$ are also CCF/BPP languages, we can reduce regular separability of VASS languages to the following problem:

Given: CCF/BPP languages K, L, regular R.
Question: Does $K \cap R|L$?

3. Approximants of $\mathbb{Z}$-VASS

Fix $n \in \mathbb{N}$. Let $\Sigma = \{a_i, \bar{a}_i \mid i \in \{1, \ldots, n\}\}$ and $\varphi\colon \Sigma^* \to \mathbb{Z}^n$ the morphism with $\varphi(a_i) = 1$ and $\varphi(\bar{a}_i) = -1$. Let $Z \subseteq \Sigma^*$ be the set of all $w \in \Sigma^*$ with $\varphi(w) = 0$.

We want to understand the regular languages R with $R \cap Z = \emptyset$. We begin with two types of such regular languages. For $k \in \mathbb{N}$, let $\mu_k\colon \mathbb{Z}^n \to (\mathbb{Z}/k\mathbb{Z})^n$ the projection modulo k. The language

$$M_k = \{w \in \Sigma^* \mid \mu_k(\varphi(w)) \neq 0\}$$

is clearly regular and satisfies $M_k \cap Z = \emptyset$.

Let $z_1, z_2, \ldots, z_m \in \mathbb{Z}$. We call the sequence *$k$-increasing* if there are indices $1 = i_0 < \ldots < i_r = m$ such that $|z_s - z_t| \leq k$ for $s, t \in [i_{\ell-1}, i_\ell]$, $\ell \in [0, r]$, and with $S_\ell = \min_{s \in [i_{\ell-1}, i_\ell]} z_s$, we have $S_1 < S_2 < \cdots < S_r$. For $y \in \mathbb{Z}^n$ and $k \in \mathbb{N}$, let $I_{y,k} \subseteq \Sigma^*$ be the set of all words $x_1 \cdots x_m$, $x_1, \ldots, x_m \in \Sigma$, such that the sequence $z_1, \ldots, z_m$ with $z_i = \langle \varphi(x_1 \cdots x_i), y\rangle$ is k-increasing and $\langle \varphi(x_1 \cdots x_m), y\rangle \neq 0$. Clearly, every language $I_{y,k}$ is regular and disjoint from Z.

Theorem 3. *Let $R \subseteq \Sigma^*$ be a regular language with $R \cap Z = \emptyset$. Then $R = \bigcup_{i=1}^{r} R_i$ where each R_i is a regular subset of either some M_k or some $I_{y,k}$.*

For the proof of theorem 3, we need two ingredients. The first is the well-known Farkas Lemma [2, Corollary 7.1d]. It tells us that non-solvability of an equation system is certified by a hyperplane separating all the left-hand sides from the right-hand side.

Lemma 4 (Rational Farkas Lemma)**.** *Let $A \in \mathbb{Q}^{n \times m}$ and $b \in \mathbb{Q}^n$. Then exactly one of the following holds:*

(1) There is a solution $x \in \mathbb{Q}^m$ to the equation $Ax = b$ with $x \geq 0$.
(2) There is a vector $y \in \mathbb{Q}^n$ with $y^t A \geq 0$ and $y^t b < 0$.

The following is the known fact that abelian groups are subgroup separable. It essentially says that non-membership in a finitely generated subgroup is certified by a morphism into a finite group.

Lemma 5. *If G is a finitely generated abelian group, H is a finitely generated subgroup, and $g \notin H$, then there is a finite abelian group F and a morphism $\psi\colon G \to F$ with $\psi(H) = 0$ and $\psi(g) \neq 0$.*

Proof. Dividing by H allows us to assume that $H = 0$, $g \neq 0$. Since G is finitely generated abelian, we have $G \cong \mathbb{Z}^n \oplus \bigoplus_{i=1}^{m} \mathbb{Z}/a_i\mathbb{Z}$. If $n = 0$, we are done. Otherwise, choose a number k that is larger than each of the torsion-free components of g. Then the projection to $(\mathbb{Z}/k\mathbb{Z})^n \oplus \bigoplus_{i=1}^{m} (\mathbb{Z}/a_i\mathbb{Z})$ maps g to an element $\neq 0$. $\square$

Proof sketch for theorem 3:

- Construct Parikh annotation for R. Use idea of [1, Prop. 9.1] to construct one where in each linear set, the periods are linearly independent. Using the Parikh annotation, we can decompose R into finite union of regular languages, each of which is equipped with a linear Parikh annotation with independent period vectors. Hence, it suffices to consider the case where there is only one linear set.
- Let $A \in \mathbb{Z}^{n \times m}$ be the matrix whose columns are the φ-images of the period vectors and let $-b \in \mathbb{Z}^n$ be the φ-image of the base vector. Since $R \cap Z = \emptyset$, we know that $Ax = b$ has no solution in $\mathbb{N}^m$.

- Suppose $Ax = b$ has no solution in $\mathbb{Z}^m$. According to lemma 5, there is a finite abelian group F and a morphism $\mu\colon \mathbb{Z}^n \to F$ such that μ maps all columns of A to 0 and $\mu(b) \neq 0$. In particular, μ maps all φ-images of words in R to $F \setminus \{0\}$. Since every morphism from $\mathbb{Z}^n$ into a finite group factorizes over some μ_k, this yields a set M_k containing R.
- Suppose $Ax = b$ has a solution in $\mathbb{Z}^m$. Then there is no solution $x \in \mathbb{Q}^m$ with $x \geq 0$: Otherwise, since the independence of the columns of A make the solution unique, this solution would belong to $\mathbb{Q}_+^m \cap \mathbb{Z}^m = \mathbb{N}^m$, which is impossible.

  We can therefore apply lemma 4 and obtain a $y \in \mathbb{Q}^n$ with $y^t A \geq 0$ and $y^t b < 0$. We can clearly choose it with $y \in \mathbb{Z}^n$. The φ-image of every word in R is of the form $-b + Ax$ with $x \in \mathbb{N}^m$. In particular, every such vector $z \in \mathbb{Z}^n$ satisfies $y^t z > 0$. Consider a loop in an automaton for R. The φ-image z' of the label of this loop must satisfy $y^t z' \geq 0$, since otherwise pumping would contradict the previous observation. Therefore, we can choose k so that R is contained in $I_{y,k}$.

Theorem 3 tells us that regular separability of languages in $\mathcal{C}$ from $\mathbb{Z}$-VASS languages reduces to the following problem:

Given: A language L from $\mathcal{C}$ and $n \in \mathbb{N}$.
Question: Does there exist k and a finite set $F \subseteq \mathbb{Z}^n$ such that $L \subseteq M_k \cup \bigcup_{y \in F} I_{y,k}$?

4. Soft-bodied $\mathbb{Z}$-VASS separability

For an alphabet X, let $X^\oplus$ denote the set of mappings $X \to \mathbb{N}$. Moreover, let $\Psi_X(\cdot)\colon X^* \to X^\oplus$ be the Parikh map. For an alphabet $Y \subseteq X$, the map $\pi_Y\colon X^* \to Y^*$ is the projection onto Y^*.

We want to solve the *$\mathbb{Z}$-VASS separation problem*:

Input: A regular language $R \subseteq X^*$ and semilinear sets $U_1, U_2 \subseteq X^\oplus$.
Question: Is there a regular language S with $R \cap \Psi_\Sigma^{-1}(U_1) \subseteq S$ and $S \cap R \cap \Psi_\Sigma^{-1}(U_1) = \emptyset$?

We show the following:

Proposition 6. *The $\mathbb{Z}$-VASS separation problem can be reduced to separability of semilinear sets by unary sets.*

The first step is to define Parikh annotations.

Definition 7. *Let $L \subseteq X^*$ be a language and $\mathcal{C}$ be a language class. A* Parikh annotation (PA) *for L in $\mathcal{C}$ is a tuple $(K, C, P, (P_c)_{c \in C}, \varphi)$, where*

- *C, P are alphabets such that X, C, P are pairwise disjoint,*
- *$K \subseteq C(X \cup P)^*$ is in $\mathcal{C}$,*
- *φ is a morphism $\varphi\colon (C \cup P)^\oplus \to X^\oplus$,*
- *P_c is a subset $P_c \subseteq P$ for each $c \in C$,*

such that

1. *$\pi_X(K) = L$ (the* projection property*),*
2. *$\varphi(\pi_{C \cup P}(w)) = \Psi(\pi_X(w))$ for each $w \in K$ (the* counting property*), and*
3. *$\Psi(\pi_{C \cup P}(K)) = \bigcup_{c \in C} c + P_c^\oplus$ (the* commutative projection property*).*

Intuitively, a Parikh annotation describes for each w in L one or more Parikh decompositions of $\Psi(w)$. The symbols in C represent constant vectors and symbols in P represent period vectors. Here, the symbols in $P_c \subseteq P$ correspond to those that can be added to the constant vector corresponding to $c \in C$. Furthermore, for each $x \in C \cup P$, $\varphi(x)$ is the vector represented by x.

For this application of Parikh annotations, we need a further property. A Parikh annotation $(K, C, P, (P_c)_{c\in C}, \varphi)$ for L is said to be *pseudo-bounded* if on each of the sets P_c, one can establish a linear order $(P_c, <)$ such that

$$cp_1^* \cdots p_n^* \subseteq \pi_{C\cup P}(K),$$

where $P_c = \{p_1, \ldots, p_n\}$ and $p_1 < \cdots < p_n$.

In other words, in a pseudo-bounded PA, when projecting to the annotation alphabet $\{c\} \cup P_c$, every description of a Parikh image appears as some word from the bounded language $cp_1^* \cdots p_n^*$. It is not hard to see that regular languages admit pseudo-bounded Parikh annotations. See [3, Lemma 9.3.5, p. 151] for a (short) proof.

Lemma 8. *Given a regular language R, one can construct a regular pseudo-bounded Parikh annotation for R.*

Reduction. We now prove proposition 6. Let $(K, C, P, (P_c)_{c\in C}, \varphi)$ be a pseudo-bounded regular Parikh annotation for R. For each $i \in \{1, 2\}$ and $c \in C$, define

$$T_{i,c} = \{\mu \in P_c^{\oplus} \mid \varphi(c + \mu) \in U_i\}.$$

Then $T_{i,c}$ is clearly Presburger-definable and hence semilinear. We establish proposition 6 by showing the following.

Lemma 9. *$R \cap \Psi_\Sigma^{-1}(U_1)$ and $R \cap \Psi_\Sigma^{-1}(U_2)$ are separable by a regular language if and only if for each $c \in C$, the sets $T_{1,c}$ and $T_{2,c}$ are separable by a unary set.*

Proof. We begin with the "if" direction. Let $S_c \subseteq P_c^{\oplus}$ be a unary separator of $T_{1,c}$ and $T_{2,c}$, meaning $T_{1,c} \subseteq S_c$ and $S_c \cap T_{2,c} = \emptyset$. Define

$$S = \{\pi_X(w) \mid \exists c \in C\colon cw \in K,\ \Psi(\pi_{P_c}(w)) \in S_c\},$$

which is clearly regular because each S_c is unary. We claim that S is a separator for $R \cap \Psi_X^{-1}(U_1)$ and $R \cap \Psi_X^{-1}(U_2)$.

Suppose $u \in R \cap \Psi_X^{-1}(U_1)$. By the projection property, there is a $c \in C$ and $cw \in K$ such that $\pi_X(w) = u$. Since $\Psi(u) \in U_1$, the counting property entails

$$\varphi(c + \Psi(\pi_{P_c}(w))) = \Psi(u) \in U_1,$$

which implies $\Psi(\pi_{P_c}(w)) \in T_{1,c} \subseteq S_c$ and thus $u \in S$.

Now assume $u \in S$. Then we can write $u = \pi_X(w)$ such that for some $c \in C$, we have $cw \in K$ and $\Psi(\pi_{P_c}(w)) \in S_c$. The latter implies $\Psi(\pi_{P_c}(w)) \notin T_{2,c}$. By definition of $T_{2,c}$, this means $\varphi(c+\Psi(\pi_{P_c}(w))) \notin U_2$ and therefore, by the counting property,

$$\Psi(u) = \varphi(c + \Psi(\pi_{P_c}(w))) \notin U_2.$$

Thus, $S \cap \Psi_X^{-1}(U_2) = \emptyset$ and S separates $R \cap \Psi_X^{-1}(U_1)$ and $R \cap \Psi_X^{-1}(U_2)$.

For the "only if" direction, suppose S is a regular language with $R \cap \Psi_X^{-1}(U_1) \subseteq S$ and $R \cap \Psi_X^{-1}(U_2) \cap S = \emptyset$. First, we modify S slightly and obtain for each $c \in C$ the regular language

$$S'_c = \{cw \in K \mid \pi_X(w) \in S,\ \pi_{P_c}(w) \in p_1^* \cdots p_n^*\},$$

where $P_c = \{p_1, \ldots, p_n\}$ such that $p_1 < \cdots < p_n$. Observe that the language $\pi_{P_c}(S'_c)$ is included in $p_1^* \cdots p_n^*$, which implies that its image $S_c = \Psi(\pi_{P_c}(S'_c)) \subseteq P_c^{\oplus}$ is unary. We claim that S_c separates $T_{1,c}$ and $T_{2,c}$.

Let $\mu \in T_{1,c}$. Then $\varphi(c+\mu) \in U_1$. Moreover, by pseudo-boundedness of the Parikh annotation, we have a word $cw \in K$ such that $\pi_{P_c}(w) \in p_1^* \cdots p_n^*$ and $\Psi(\pi_{P_c}(w)) = \mu$. Since $\pi_X(cw) \in R$ (projection property) and $\Psi(\pi_X(cw)) = \varphi(c+\mu) \in U_1$, we have $\pi_X(cw) \in S$ and hence $cw \in S'_c$. Thus $\mu = \Psi(\pi_{P_c}(w)) \in S_c$.

Now let $\mu \in S_c$, which means there is a $cw \in K$ with $\pi_X(cw) \in S$ and $\Psi(\pi_{P_c}(w)) = \mu$. Together with $\pi_X(cw) \in R$, the fact $\pi_X(cw) \in S$ tells us that $\Psi(\pi_X(cw)) \notin U_2$ and thus

$$\varphi(c+\mu) = \Psi(\pi_X(cw)) \notin U_2,$$

hence $\mu \notin T_{2,c}$ by definition. $\square$

References

[1] Samuel Eilenberg and Marcel-Paul Schützenberger. "Rational sets in commutative monoids". In: *Journal of Algebra* 13.2 (1969), pp. 173–191. DOI: 10.1016/0021-8693(69)90070-2.

[2] Alexander Schrijver. *Theory of linear and integer programming*. John Wiley & Sons, 1986.

[3] Georg Zetzsche. "Monoids as Storage Mechanisms". PhD thesis. Technische Universität Kaiserslautern, 2016.