Semilinear Sets

We have discussed linear integer systems.
As we have shown, the solution spaces of these sets are of the form $B+P^*$.
To be able to reason about the properties such sets, we need to study a more general object: Semilinear Sets.

Definition: We call $L \subseteq \mathbb{N}^d$ a linear, if there is a $b \in \mathbb{N}^d$ and finite $P \subseteq \mathbb{N}^d$ with $L = b + P^*$.

A set $S \subseteq \mathbb{N}^d$ is called semilinear (SL), if $S = \bigcup_{i \leq k} L_i$ for some linear $L_1, \dots, L_k \subseteq \mathbb{N}^d$.

Our previous theorem implies

Corollary: The solution spaces of linear integer systems are semilinear.

Semilinear sets are important beyond the study of linear integer systems.
An important application is the Parikh Image of regular and context free languages.
Let Σ be an alphabet.
For a word $w \in \Sigma^*$, we define $\Psi(w) \in \mathbb{N}^\Sigma$ as the vector that counts the appearances of each letter

e.g. $\Psi(abcca) = \begin{pmatrix} 2 \\ 1 \\ 1 \end{pmatrix}$ for $\Sigma = \{a,b,c\}$

For a language $L \subseteq \Sigma^*$, we write $\Psi(L) = \{\Psi(w) \mid w \in L\} \subseteq \Sigma^*$.

Then the following holds:

Parikh's Theorem: Let L be a context-free language. Then $\Psi(L)$ is semilinear.

Now, we will show the closure properties of SL sets, namely

Theorem: Let $S, T \subseteq \mathbb{N}^d$ be semilinear sets and $A \in \mathbb{N}^{\ell \times d}$. Then so are

(Union)	$S \cup T$
(Projection)	$\pi(S)$
(Product)	$S \times T$ (where $\pi : \mathbb{N}^d \to \mathbb{N}^k$, $\pi(x) = (x[i_1], \dots, x[i_k])$)
(Linear Map)	$\tau(S)$ ← For all s,t, $\tau(s+t) = \tau(s) + \tau(t)$.
(Intersection)	$S \cap T$
(Complement)	$\mathbb{N}^d \setminus S$

It is easy to see closure under union.

Proof (Union): $S = \bigcup_{i \leq k} L_i$ and $T = \bigcup_{i \leq m} K_i$ where $L_1, \dots, L_k$ and $K_1, \dots, K_m$ are linear.

Then $S \cup T = \bigcup_{i \leq k} L_i \cup \bigcup_{i \leq m} K_i$ is also semilinear. □

For projection, we only need to show that $\pi(L)$ is linear for linear L, since

$\pi(\bigcup_{i \leq k} L_i) = \bigcup_{i \leq k} \pi(L_i)$.

Lemma: Let $L = b + P^*$ be linear. Then so is $\pi(L)$.

Proof: It suffices to apply the projection to b and elements in P.

We claim $\pi(L) = \pi(b) + \pi(P)^*$.

This holds since $\pi(b) + \pi(p_1) + \dots + \pi(p_k) = \pi(b + p_1 + \dots + p_k)$ for $p_1, \dots, p_k \in P$.

For product, it also suffices to reason about linear sets since $\bigcup_{i \leq k} L_i \times \bigcup_{j \leq m} K_j = \bigcup_{\substack{i \leq k \\ j \leq m}} L_i \times K_j$

Lemma: If $L = b + P^*$ and $K = c + R^*$ for finite $P, R \subseteq \mathbb{N}^d$ and $b, c \in \mathbb{N}^d$, we have

$$L \times K = (b, c) + ((P \cup \{0\}) \times (R \cup \{0\}))^*$$

Proof: $\subseteq$: Let $(x, y) \in L \times K$. Then $x = b + p_1 + \dots + p_k$ for $p_1, \dots, p_k \in P$ and $y = c + r_1 + \dots + r_m$ for $r_1, \dots, r_m \in R$.

Then $(x, y) = (b, c) + \underbrace{(p_1, 0) + \dots + (p_k, 0) + (0, r_1) + \dots + (0, r_m)}_{\in ((P \cup \{0\}) \times (R \cup \{0\}))^*}$

$\supseteq$: Let $(x, y) = (b, c) + (p_1, r_1) + \dots + (p_k, r_k)$ where $(p_i, r_i) \in ((P \cup \{0\}) \times (R \cup \{0\}))$ for all $i \leq k$.

Then $x = b + p_1 + \dots + p_k$ for $p_1, \dots, p_k \in P \cup \{0\}$ and

$y = c + r_1 + \dots + r_k$ for $r_1, \dots, r_k \in R \cup \{0\}$.

Clearly, $p_1 + \dots + p_k \in P^*$ and $r_1 + \dots + r_k \in R^*$. □

Now, we show that linear sets are closed under linear maps A, which implies closure for SL sets.

Lemma: Let $L = b + P^*$ for finite $P \subseteq \mathbb{N}^d$, $b \in \mathbb{N}^d$, and $\tau : \mathbb{N}^d \to \mathbb{N}^\ell$ a linear map. Then

$$\tau(L) = \tau(b) + \tau(P)^*$$

Proof: It is easy to see that $\tau(x) = A \cdot x$ for some $A \in \mathbb{N}^{\ell \times d}$. Then,

$$\begin{aligned} A \cdot L = A \cdot (b + P^*) = A \cdot b + A \cdot P^* &= A \cdot b + \{A \cdot (p_1 + \dots + p_k) \mid p_1, \dots, p_k \in P\} \\ &= A \cdot b + \{A \cdot p_1 + \dots + A \cdot p_k \mid p_1, \dots, p_k \in P\} \\ &= A \cdot b + \{p_1' + \dots + p_k' \mid p_1', \dots, p_k' \in A \cdot P\} = A \cdot b + (A \cdot P)^* = \tau(b) + \tau(P)^*. \end{aligned}$$

□

This also implies that linear (and thus SL) set are closed under addition.

Lemma: Let $L, K \subseteq \mathbb{N}^d$ be linear. Then so is $L+K$.

Proof: Let $A = \begin{bmatrix} I_d & 0 \\ 0 & I_d \end{bmatrix}$. Then $K \times L$ and thus $A \cdot (K \times L) = \{A \cdot \binom{x}{y} \mid x \in K, y \in L\}$

$= \{x+y \mid x \in K, y \in L\}$

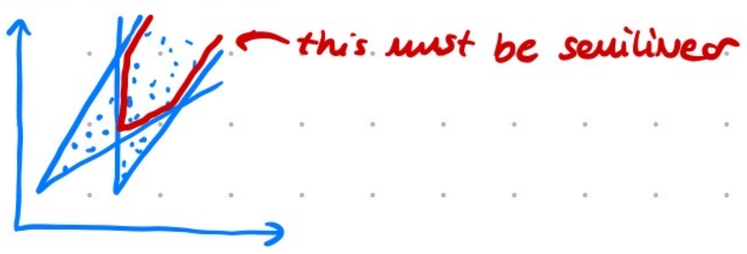
is linear. □

We show that the intersection of two linear sets is a SL set, which shows that SL sets are closed under intersection, since $\cap$ distributes over $\cup$.

$$\bigcup_{i \leq k} L_i \cap \bigcup_{i \leq m} K_i = \bigcup_{\substack{i \leq k \\ j \leq m}} L_i \cap K_j$$

Lemma: Let $L = b + P^*$ and $K = c + R^*$ for finite $P, R \subseteq \mathbb{N}^d$, and $b, c \in \mathbb{N}^d$.
Then $L+K$ is semilinear

Picture of the task



Proof: Let $P = [p_1, \dots, p_k]$ and $R = [r_1, \dots, r_m]$

We have

$$L \cap K = \{x \in \mathbb{N}^d \mid \exists u_1, \dots, u_k \in \mathbb{N}. \exists v_1, \dots, v_m \in \mathbb{N}. x = b + \sum_{i \leq k} u_i \cdot p_i = c + \sum_{i \leq m} v_i \cdot r_i\}$$

We can express which u's and v's have this property by an eq. system S

$(u,v) \in \mathbb{N}^{k+m}$ s.t.

$$[p_1 \dots p_k] \cdot u - [r_1 \dots r_m] \cdot v = c - b$$

Then $L \cap K = \{[p_1 \dots p_k] \cdot u \mid (u,v) \in sol(S)\}$

$= A \cdot sol(S)$ where $A = [p_1 \dots p_k \underbrace{0 \dots 0}_{m \text{ times}}]$.

Since $sol(S)$ is semilinear and A a linear map, so is $A \cdot sol(S)$ SL.

The final closure property needs more work.
We do this work here, and postpone the result to the next lecture.

Lemma: Every linear set $L = b + P^*$ can be written as $L = \bigcup_{i \leq k} b_i + P_i^*$ where P_i are linearly independent or $P_i = [0]$.

Proof: The proof is by induction on $|P|$.

Base Case $|P| = 1$: Then $L = b + [p]^*$. We are done.

Inductive Case $|P| = k+1$: If P is linearly independent, we are done.
Let P be linearly dependent.

Let $P = \{p_1, \dots, p_{k+1}\}$. Then there are $u_1, \dots, u_{k+1} \in \mathbb{Z}$ with

$$\sum_{i \le k+1} u_i \cdot p_i = 0,$$

We can move the negative coefficients to the side and get

$$\sum_{i \in I} u_i \cdot p_i = \sum_{i \notin I} u_i \cdot p_i \quad \text{for } I = \{i \le k+1 \mid u_i \ge 0\}$$

By relabeling, we get $a_1, \dots, a_{k+1} \in \mathbb{N}$ and an $1 \le m < k+1$ with one $j > m$ where $a_j \ge 1$ and

$$\sum_{i \le m} a_i \cdot p_i = \sum_{i > m} a_i \cdot p_i$$

Idea: Use this to remove p_j periods for $j > m$.
When there aren't enough periods to remove, a period p_j is taken boundedly. Capture this in fin. many base vectors

For each $j > m$, we define $C_j = \{b + t_j \cdot p_j \mid 0 \le t_j < a_j\}$ if $a_j > 0$
$C_j = \{b\}$ if $a_j = 0$
$P_j = P \setminus \{p_j\}$

We claim $L = \bigcup_{j > m} C_j + P_j^*$. Since $|P_j| < |P|$, the induction hypothesis decomposes each $c + P_j$ for $c \in C_j$ and $j > m$ into a union of lin. sets with lin. indep. periods, we will have shown our claim.

The $\supseteq$-direction is easy, since $C_j \subseteq L$ and $P_j \subseteq P$.
For the $\subseteq$-direction, let $y \in L$. Then, $y = b + \sum_{i \le k+1} t_i \cdot p_i$ for $t_1, \dots, t_{k+1} \in \mathbb{N}$. We have three cases.

Case 1, $\exists j > m.\ t_j = 0$: we get $y = b + \sum_{i \le k+1} t_i \cdot p_i \in b + P_j^* = C_j + P_j^*$.

Case 2, $\exists j > m.\ t_j < a_j$: Then $y = \underbrace{b + t_j \cdot p_j}_{\in C_j} + \underbrace{\sum_{i \le m} t_i \cdot p_i + \sum_{\substack{i > m \\ i \ne j}} t_i \cdot p_i}_{P_j^*} \in C_j + P_j^*$

Case 3, $\forall j > m.\ t_j \ge a_j$: There is an $\ell \in \mathbb{N}$ such that $\forall j > m.\ (t_j - \ell \cdot a_j) \ge 0$ and $\exists j > m.\ \underbrace{(t_j - (\ell+1) \cdot a_j)}_{t_j - \ell \cdot a_j < a_j} < 0$.
To see that such an ℓ must exist, consider that only boundedly many ℓ may satisfy the first part, since $a_j \ne 0$ for some $j > m$.
Then, for the largest such ℓ, $\ell+1$ does not satisfy the condition, which implies that the second part also holds for ℓ.

Now, $y = b + (t_j - \ell \cdot a_j) \cdot p_j + \sum_{i \le m} t_i \cdot p_i + \sum_{\substack{i > m \\ i \ne j}} \ell \cdot a_i \cdot p_i + \sum_{\substack{i > m \\ i \ne j}} (t_i - \ell \cdot a_i) \cdot p_i$ (rewrite)

$$= b + (t_j - \ell \cdot a_j) \cdot p_j + \sum_{i \le m} t_i \cdot p_i + \sum_{\substack{i > m \\ i \ne j}} \ell \cdot a_i \cdot p_i + \sum_{\substack{i > m \\ i \ne j}} (t_i - \ell \cdot a_i) \cdot p_i$$

$$= \underbrace{b + (t_j - \ell \cdot a_j) \cdot p_j}_{\in C_j} + \underbrace{\sum_{i \le m} (t_i + \ell \cdot a_i) \cdot p_i + \sum_{\substack{i > m \\ i \ne j}} (t_i - \ell \cdot a_i) \cdot p_i}_{P_j^*} \in C_j + P_j^*$$

($t_j - \ell \cdot a_j < a_j$)

Now we show that $\mathbb{N}^d \setminus L$, where L has lin. independent periods, is semilinear.
This shows that $\mathbb{N}^d \setminus S$ is semilinear for any semilinear $S \subseteq \mathbb{N}^d$, since

$$\mathbb{N}^d \setminus S = \mathbb{N}^d \setminus \bigcup_{i \le k} L_i = \mathbb{N}^d \setminus \bigcup_{i \le m} L_i' = \bigcap_{i \le m} \mathbb{N}^d \setminus L_i'$$

for linear $L_1, \dots, L_k$ and period-independent linear $L_1', \dots, L_m'$ and semilinear sets are closed under intersection.

Lemma: $\mathbb{N}^d \setminus L$ is semilinear for $L = b + P^* \subseteq \mathbb{N}^d$ with linearly independent P.

Idea: Use linear independence to construct an invertible matrix A where $A^{-1} \cdot y$ gives the only combination of coefficients for $P = [p_1, \dots, p_k]$ vectors that yield y. We can describe whether $A^{-1}y$ values are bad ($y \notin L$) using a disjunction of lin. int. systems $\Rightarrow$ semilinear.

Proof: Let $P = [p_1, \dots, p_k]$.

Let $\{r_{k+1}, \dots, r_d\} \subseteq [e_1, \dots, e_d\}$ of unit vectors where $P \cup \{r_{k+1}, \dots, r_d]$ is lin. independent.

Let $A = \left[p_1 \dots p_k \, r_{k+1} \dots r_d\right]$. Then A is invertible, so there is $A^{-1} \in \mathbb{Q}^{d \times d}$ such that $A \cdot u = y$ iff $u = A^{-1} \cdot y$.

Clearly, $y \in L$ iff $A^{-1} \cdot (y-b)[i] \in \mathbb{N}$ for all $i \leq k$

and $A^{-1} \cdot (y-b)[i] = 0$ for all $i > k$

because $(y-b)$ can only be obtained by adding $A_{-,i}$ (i-th column), $A^{-1} \cdot y[i]$ times.

Since $A \in \mathbb{Q}^{d \times d}$, there is a $k \in \mathbb{N}$ with $k \cdot A^{-1} \in \mathbb{Z}^{d \times d}$.

Let $H = k \cdot A^{-1} \in \mathbb{Z}^d$.

Then, $y \in L$ iff $H \cdot (y-b)[i] \geq 0$ and $H \cdot (y-b)[i] \equiv 0 \pmod k$ for all $i \leq k$

and $H \cdot (y-b)[i] = 0$ for all $i > k$.

Then $y \notin L$ (for $y \in \mathbb{N}^d$) iff one of the following holds for some $i \leq d$.

- $i \leq k$ and $H \cdot (y-b)[i] \leq -1$ or $H \cdot (y-b)[i] \not\equiv 0 \pmod k$
- $i > k$ and $H \cdot (y-b)[i] \leq -1$ or $H \cdot (y-b)[i] \geq 1$

Thus, $y \notin L$ if y fulfills

$$\bigvee_{i \leq k} \Big((H_i \cdot (y-b) \leq -1) \vee \bigvee_{0 < u < k} (\exists t \in \mathbb{N}.\ H_i \cdot (y-b) - t \cdot k = u) \Big) \vee \bigvee_{i > k} \begin{pmatrix} (H_i \cdot (y-b) \geq 1) \vee \\ (H_i \cdot (y-b) \leq -1) \end{pmatrix}$$

The y that fulfills $H_i \cdot (y-b) \leq -1$, $\exists t \in \mathbb{N}.\ H_i \cdot (y-b) - t \cdot k = u$, or $H_i \cdot (y-b) \geq 1$ form semilinear sets, since they can be expressed as (the proj. of) solutions to linear integer systems.

□