Lemma 1. Let $P = \{p_1, \ldots, p_n\} \subset \mathbb{R}^d$ be a finite set of points in $d$-dimensional space. If $n > d$ then we have, for some coefficients $\mu_1, \ldots, \mu_n$,

$$0 = \sum_{i=1}^{n} \mu_i p_i \text{ with } \mu_1, \ldots, \mu_n \text{ not all zero}, \quad (1)$$

and if $n > d + 1$ we can have (1) under the additional condition that

$$\sum_{i=1}^{n} \mu_i = 0. \quad (2)$$

In the latter case, some of $\mu_1, \ldots, \mu_n$ are positive and some are negative.

Proof. The first part of the lemma follows from the fact that every set of $d + 1$ or more points in a $d$-dimensional vector space is linearly dependent. The second part follows from the observation that $\{p_2 - p_1, p_3 - p_1, \ldots, p_n - p_1\}$ is linearly dependent, thus

$$0 = \sum_{i=2}^{n} \mu_i (p_i - p_1) \text{ with } \mu_2, \ldots, \mu_n \text{ not all zero.}$$

By defining $\mu_1 = -\sum_{i=2}^{n} \mu_i$, both (1) and (2) then hold. The only way that (2) can hold with all non-positive or all non-negative terms would be if all terms are zero. 

\[\square\]
Now let $P$ be any (not necessarily finite) set of points in $\mathbb{R}^d$. The **convex hull** of $P$, denoted $\text{Conv}(P)$, is the set of all convex combinations of points of $P$. In other words, a point $x \in \mathbb{R}^d$ is in $\text{Conv}(P)$ if and only if, for some positive integer $n$, and some set $\{p_1, \ldots, p_n\} \subseteq P$, and some set of coefficients $\{\alpha_1, \ldots, \alpha_n\}$ with $\alpha_i \geq 0$ for all $i \in \{1, \ldots, n\}$ and $\sum_{i=1}^n \alpha_i = 1$, we have

$$x = \sum_{i=1}^n \alpha_i p_i.$$  

**Theorem** (Carathéodory). For $P \subseteq \mathbb{R}^d$, if $x \in \text{Conv}(P)$ then $x \in \text{Conv}(P')$ for some subset $P'$ of $P$ of cardinality at most $d + 1$.

**Proof.** Let $x$ be a point of $\text{Conv}(P)$, so that for some positive integer $n$

$$x = \sum_{i=1}^n \alpha_i p_i$$

with, for all $i \in \{1, \ldots, n\}$, $p_i \in P$, $\alpha_i \geq 0$, and $\sum_{i=1}^n \alpha_i = 1$. If $n \leq d + 1$ there is nothing to prove. Otherwise, $n > d + 1$, so by Lemma 1, we have for scalars $\mu_1, \ldots, \mu_n$ (some of which are positive) that $0 = \sum_{i=1}^n \mu_i p_i$, with $\sum_{i=1}^n \mu_i = 0$.

Now, for any real number $\lambda$, we have

$$x = \sum_{i=1}^n \alpha_i p_i - \lambda \sum_{i=1}^n \mu_i p_i = \sum_{i=1}^n (\alpha_i - \lambda \mu_i)p_i. \quad (3)$$

Note that

$$\sum_{i=1}^n (\alpha_i - \lambda \mu_i) = \sum_{i=1}^n \alpha_i - \lambda \sum_{i=1}^n \mu_i = \sum_{i=1}^n \alpha_i - \lambda \cdot 0 = 1,$$

i.e., the coefficients in the linear combination (3) sum to one. We will now select $\lambda$ so that one of these coefficients becomes zero, while the remaining coefficients are positive, making (3) a convex combination of $n - 1$ points of $P$.

Let $J = \{j \in \{1, \ldots, n\} : \mu_j > 0\}$ (and note that $J$ is not empty). Choose $j^* \in J$ so that $\alpha_{j^*} / \mu_{j^*} \leq \alpha_j / \mu_j$ for all $j \in J$, and let $\lambda = \alpha_{j^*} / \mu_{j^*}$. With this choice of $\lambda$, we have

$$\alpha_i - \lambda \mu_i \geq 0$$
for all \( i \in \{1, \ldots, n\} \). Indeed if \( i \in J \), then \( \mu_i > 0 \) and
\[
\alpha_i - \lambda \mu_i = \mu_i (\alpha_i/\mu_i - \lambda) \geq 0,
\]
while if \( i \not\in J \) then \( \mu_i \leq 0 \), and since \( \lambda \geq 0 \), we have
\[
\alpha_i - \lambda \mu_i \geq \alpha_i \geq 0.
\]
Finally observe that \( \alpha_{j^*} - \lambda \mu_{j^*} = 0 \), so
\[
x = \sum_{i=1}^{j^*-1} (\alpha_i - \lambda \mu_i) p_i + \sum_{i=j^*+1}^{n} (\alpha_i - \lambda \mu_i) p_i,
\]
which expresses \( x \) as a convex combination of the \( n - 1 \) elements of the set \( \{p_1, \ldots, p_n\} \setminus \{p_{j^*}\} \). This process can be repeated as long as \( n > d + 1 \), until \( x \) is represented as a convex combination of \( d + 1 \) elements of \( P \).

**Theorem (Radon).** Every set of \( d + 2 \) points in \( \mathbb{R}^d \) can be partitioned into two sets \( P_1 \) and \( P_2 \) such that \( \text{Conv}(P_1) \cap \text{Conv}(P_2) \neq \emptyset \).

**Proof.** Let \( n = d + 2 \) and let \( P = \{p_1, \ldots, p_n\} \subseteq \mathbb{R}^d \). By Lemma 1 we have for scalars \( \mu_1, \ldots, \mu_n \) (some of which are positive and some of which are non-positive) \( 0 = \sum_{i=1}^{n} \mu_i p_i \) with \( \sum_{i=1}^{n} \mu_i = 0 \).

Let \( J = \{j \in \{1, \ldots, n\} : \mu_j > 0\} \) and let \( I = \{1, \ldots, n\} \setminus J \). Neither \( J \) nor \( I \) is empty. Let \( A = \sum_{j \in J} \mu_j \) and note that \( A > 0 \). Then \( \sum_{i \in I} \mu_i = -A \). Finally, let \( P_1 = \{p_j : j \in J\} \) and \( P_2 = \{p_i : i \in I\} \).

The convex hulls \( \text{Conv}(P_1) \) and \( \text{Conv}(P_2) \) contain the point
\[
p = \sum_{j \in J} \frac{\mu_j}{A} x_j = \sum_{i \in I} \frac{-\mu_i}{A} x_i
\]
as the first sum is a convex combination of points from \( P_1 \) and the second sum is a convex combination of points from \( P_2 \). Thus \( \text{Conv}(P_1) \cap \text{Conv}(P_2) \neq \emptyset \). \( \square \)

If \( P = \{p_1, \ldots, p_{d+2}\} \subseteq \mathbb{R}^d \) is partitioned into subsets \( P_1 \) and \( P_2 \) then a point \( p \in \text{Conv}(P_1) \cap \text{Conv}(P_2) \) is called a Radon point of \( P \). For example, the (only) Radon point of \( P = \{p_1, p_2, p_3\} \subseteq \mathbb{R} \) is the median of \( P \).

Recall that a subset \( X \) of \( \mathbb{R}^d \) is convex if it contains the convex hull of its subsets, i.e., \( X \) is convex if \( \text{Conv}(Y) \subseteq X \) for every \( Y \subseteq X \). Also recall that the intersection of two convex sets is again convex.

3
Theorem (Helly). Let $X_1, \ldots, X_n$ be a collection of convex subsets of $\mathbb{R}^d$, with $n > d + 1$. If the intersection of every $d + 1$ of these sets is nonempty, then these subsets have a point in common, i.e.,

$$\bigcap_{i=1}^n X_i \neq \emptyset.$$  

Proof. We proceed by induction on $n$.

Consider the base case, $n = d + 2$. Then the intersection of any $n - 1$ of the subsets is nonempty. For each $i \in \{1, \ldots, n\}$, let $x_i$ be a point in common to all the subsets except (possibly) $X_i$. If $x_1, \ldots, x_n$ are not all distinct, then a repeated element is a point in common to all subsets. Otherwise, according to Radon’s Theorem, the set $\{x_1, \ldots, x_n\}$ can be partitioned into two subsets $P_1$ and $P_2$ such that $\text{Conv}(P_1)$ and $\text{Conv}(P_2)$ have a point $p$ in common. For each $i \in \{1, \ldots, n\}$, either $P_1 \subset X_i$ or $P_2 \subset X_i$, thus, since $X_i$ is convex, either $\text{Conv}(P_1) \subset X_i$ or $\text{Conv}(P_2) \subset X_i$. In either case $p \in X_i$, which establishes that $p$ is a point in common to all subsets, and therefore the theorem is true in the base case.

Suppose the induction hypothesis is true for some $n \geq d + 2$, and let $X_1, \ldots, X_{n+1}$ be a collection of convex subsets of $\mathbb{R}^d$ with the property that the intersection of every $d + 1$ of them is nonempty. Note that $X_n \cap X_{n+1}$ is a convex subset of $\mathbb{R}^d$. Consider the collection $X_1, \ldots, X_{n-1}, X_n \cap X_{n+1}$ of $n$ convex subsets of $\mathbb{R}^d$. Take any $d + 1$ of the sets $\{X_1, \ldots, X_n\}$; then, by assumption they have a point in common. Otherwise, take $d$ of the sets $\{X_1, \ldots, X_n\}$ together with $X_n$ and $X_{n+1}$. This is a collection of $d + 2$ convex subsets of $\mathbb{R}^d$ to which the base case case applies, so they also must have a point in common. Thus the intersection of every $d + 1$ of $X_1, \ldots, X_{n-1}, X_n \cap X_{n+1}$ is nonempty. By the induction hypothesis, these sets have a point $p$ in common, i.e., $X_1 \cap \cdots \cap X_n \cap X_{n+1}$ is nonempty, which shows that the induction hypothesis is true for $n + 1$.

Since the induction hypothesis is true for $n = d + 2$, and the truth of the induction hypothesis for $n \geq d + 2$ implies its truth for $n + 1$, by induction the hypothesis is true for all $n \geq d + 2$. \qed