• One more example of how probability can help in high dim's computational geometry

Def: A set $T \subseteq \mathbb{R}^n$ is **convex** if $\forall x, y \in T$, $[x, y] \subseteq T$.

Ex: 
- Convex
- Convex
- Non-convex
- Non-convex

• How can one transform nonconvex set $\rightarrow$ convex?

Def: The **convex hull of a set** $T \subseteq \mathbb{R}^n$, denoted $\text{conv}(T)$, is the smallest convex set that contains $T$.

Ex: 
(a) $\text{conv}([a, b]) = [a, b]$
(b) $A \rightarrow B$
(c) $A \rightarrow B$
(d) $A \rightarrow B$

$T$ = cities in Ukraine

\[\text{conv}(T)\]
FACT (CONVEX COMBINATION)

\[ \forall z \in \text{conv}(T) \text{ can be expressed as} \]

\[ z = \sum_{i=1}^{m} \lambda_i z_i, \text{ where } \lambda_i \geq 0, \sum_{i=1}^{m} \lambda_i = 1, \ z_i \in T \quad (*) \]

*PROOF*: DIY (do it yourself).

(*) is called a "convex combination". It is a linear combination, similar to a basis expansion of \( z \), but non-unique (if \( m > n \)).

\[ \begin{align*}
  \text{Ex:} & \quad z = \frac{1}{4} z_1 + \frac{1}{4} z_2 + \frac{1}{4} z_3 + \frac{1}{4} z_4 \\
  \text{Ex:} & \quad z = 0.4 z_1 + 0.3 z_2 + 0.2 z_3 + 0.1 z_4
\end{align*} \]

\[ \text{conv} \{ z_1, z_2, z_3, z_4 \} \]

Q: Computing convex hull: how large is \( m \)?

Caratheodory Theorem: \( \forall \) point in \( \text{conv}(T) \) can be expressed as a convex combination of \( \leq n+1 \) points from \( T \).

Ex: Kyiv \( \in \text{conv} \{ \text{Lutsk, Chernihiv, Simpheropol} \} \) \( n=2 \).
Remark: $n+1$ is unimprovable (e.g. in the Kyiv example above).

But: if we allow to approximate $x$, then a dramatic improvement!

Theorem (Approx. Carathéodory) Let $T \subset \mathbb{R}^n$, $\text{diam}(T) \leq 1$.

Then \( \forall x \in \text{conv}(T), \forall k \in \mathbb{N} \quad \exists \ x_1, \ldots, x_k \in T: \)

\[ \left\| x - \frac{1}{k} \sum_{i=1}^{k} x_i \right\|_2 \leq \frac{1}{\sqrt{2k}} \]

where Euclidean norm in $\mathbb{R}^n$.

Remark: Dimension-free!

Does not depend on $n$ or geometry of $T$.

Proof by a probabilistic method: the "empirical method of Maurey"

Will use standard facts of probability:

1. Def of expectation of a discrete r.v. $X$:
   
   If $X$ takes values $x_i$ with prob. $p_i$,
   
   \[ E[X] \overset{\text{def}}{=} \sum_i p_i x_i \]

   (analogous to continuous case)

   \[ \text{where } E[X] = \int_{-\infty}^{\infty} x \, p(x) \, dx \]

2. \( \text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2 \)

   \( = \frac{1}{2} E[(X - X')]^2 \) (linearity & independence)

   \( \text{where } X' \text{ is an independent copy of } X \)

   (i.e. $X, X'$ are independent and have the same distribution)

   Proof of (2):

   \[ E[(X - X')^2] = E[X^2 - 2XX' + (X')^2] \]

   \[ = E[X^2] - 2E[X]E[X'] + \left( E[X']^2 \right) \]

   \[ = \frac{1}{2} E[(X - X')]^2 \]

   (same distribution)

   \( \text{where } X' \text{ is an independent copy of } X \)

3. Extend (2) for random vectors $X$ in $\mathbb{R}^n$:

   \[ E\|X - E[X]\|_2^2 = \frac{1}{2} E\|X - X'\|_2^2 \]

   where $X'$ is an independent copy of $X$ (HW2)
Proof of Approximate C.T.

1. Fix $x \in \text{conv}(T)$, express it as a convex combination

   \[ x = \sum_{i=1}^{m} \lambda_i z_i, \quad \lambda_i \geq 0, \quad \sum_{i=1}^{m} \lambda_i = 1, \quad z_i \in T. \]

   - potentially large

2. Interpret $\lambda_i$ as probabilities, the sum as expectation

   let r.vector $Z$ take value $z_i$ with prob. $\lambda_i$

   \[ \Rightarrow \quad x = E[Z] \]

3. Consider indep. copies $Z_1, Z_2, \ldots$ of $Z$.

   SLLN: $\frac{1}{k} \sum_{i=1}^{k} Z_i \rightarrow E[Z] = x \quad \text{a.s.} \]

   \[ P(E) \]

4. Error:

   \[ E \left\| x - \frac{1}{k} \sum_{i=1}^{k} Z_i \right\|^2 = E \left\| \frac{1}{k} \sum_{i=1}^{k} (Z_i - \frac{1}{k} E[Z]) \right\|^2 \]

   \[ x = (x_1, \ldots, x_n) \in \mathbb{R}^n \quad \|x\|^2 = x_1^2 + \ldots + x_n^2 \]

   \[ = \frac{1}{k^2} \sum_{i=1}^{k} E \left\| Z_i - E[Z] \right\|^2 = \frac{1}{k^2} \sum_{i=1}^{k} \frac{\|Z_i - E[Z]\|^2}{E\|Z - E[Z]\|^2} \]

   (variance of sum = sum of variances, for vector)

   \[ = \frac{1}{k} E \left\| Z - E[Z] \right\|^2 = \frac{1}{2k} E \left\| Z - Z' \right\|^2 \]

   (fact 3 previous page)

   \[ \leq \frac{1}{2k} \]

   \[ \Rightarrow \exists \text{ realization of the r.v.'s } Z_1, \ldots, Z_k \text{ s.t.} \]

   \[ \| x - \frac{1}{k} \sum_{i=1}^{k} Z_i \|_2 \leq \frac{1}{2k}. \quad \text{Since } z_i \in T, \quad \text{QED} \]

   \[ -4 - \]
Application of Approx. Carathéodory Thm:

Cocktail Problem: You are given $N$ glasses with different cocktails, each made by mixing $n$ ingredients in different proportions. Make a glass of cocktail with given proportions $p_1, \ldots, p_n$.

Equivalently, we need to find $\lambda_1$ (portion of glass 1), $\lambda_2$ (portion of glass 2), \ldots such that:

$$p = \sum_{i=1}^{N} \lambda_i z_i, \quad \lambda_i \geq 0, \quad \sum_{i=1}^{N} \lambda_i = 1$$

proportions need \text{ to be nonnegative }

\begin{itemize}
  \item i.e., we need to express $p$ as a convex combination of vectors $z_1, \ldots, z_N$
  \item Convex program finds a solution in polynomial time.
  \item Approx. Carathéodory Thm transforms it into an approximate solution with few glasses, mixed in equal proportions; an independent of $n, N$.\footnotetext{5}
\end{itemize}
Relevance of cocktail problem

(a) (Portfolio building)

ingredients = stocks
glasses of cocktails = mutual funds
empty glass = portfolio

Problem: create a new mutual fund with a
given combination of stocks
by combining the mutual funds that are available
on the market.

Solution: as above - a fast randomized algorithm builds
a portfolio with few mutual funds.

(6) (Factor analysis)

$z_1 \ldots z_N$ = a dictionary of factors
that need to explain $z$ = behavior (consumer, animal, etc)

Sol: behavior is explained by few factors.

$z = \Sigma \lambda_i z_i \Rightarrow$ factor 1 explains $\lambda_i \%$ of behavior

$\Rightarrow$ A PARSIMONIOUS MODEL.