Lecture 6: The Maurey Argument and Approximate Caratheodory

Isabella Zhu

25 February 2025

§1 TLDR

Caratheodory, approximate Caratheodory, application to ϵ-net for polytopes, and algorithmic proof of approx Caratheodory.

§2 Caratheodory

Let $X_1, \ldots, X_n \in \mathbb{R}^d$ be given vectors. Let $C = \{\sum_{j=1}^n \lambda_j x_j, \lambda_j \geq 0, \sum \lambda_j = 1\}$ be the convex hull. For any $z \in C$, we have that we can take most λ_j equal to 0, i.e. very sparse, if $n >> d$.

Theorem 2.1

(Caratheodory) For every $z \in C$, there exists $S \subset [n]$ such that $|S| \leq d + 1$ and $Z = \sum_{j \in S} \lambda_j x_j$. This is independent of n.

Proof. Assume $n \geq d + 2$. Look at $x_2 - x_1$, $x_3 - x_1$, etc. There are $n - 1$ of these, and $n - 1 \geq d + 1$. Since we're in dimension d, these vectors are linearly dependent. So, there exists $\mu_2, \ldots \mu_n$ not all equal to 0 such that

$$\sum_{j=2}^{n} \mu_j (x_j - x_1) = 0.$$

Define $\mu_1 = -(\mu_2 + \ldots + \mu_n)$. Then,

$$\sum_{j=1}^{n} \mu_j x_j = \sum_{j=2}^{n} \mu_j (x_j - x_1) = 0.$$

Consider

$$z = \sum (\lambda_j - \beta \mu_j) x_j = \sum \lambda_j \mu_j.$$

Take $\beta = \min_{j, \mu_j > 0} \lambda_j / \mu_j$. Let the argmin be $\bar{j}$. Define $\lambda'_j = \lambda_j - \beta \mu_j$. Some analysis reveals that $\lambda_j' \geq 0$. We have

$$\sum_{j=1}^{n} \lambda_j' = \sum \lambda_j = 1.$$

So now we have gone from n to $n-1$. We can induct down now.

Remark 2.2. Note that Caratheodory is exact. But there's an approximate version, which means if we're willing to get not exactly z, then we can do even better.

Theorem 2.3

(Approximate Caratheodory) Assume $|x_j|_2 \leq 1$. Then, for every $z \in C$ and every $k \geq 0$, there exists a subset $S \subset [n]$ such that $|S| = k$ and

$$\left| z - \sum_{j \in S} \lambda_j x_j \right|_2 \leq \frac{1}{\sqrt{k}}$$

for $\lambda_j \geq 0$ and $\sum \lambda_j = 1$. In fact, we can take $\lambda_j \in \{\frac{i}{n} \mid i = 1, \dots n\}$.

Remark 2.4. This is only useful if $k \leq d$. This is also independent of n.

Proof (Maurey). If $z \in C$, then $z = \sum_{j=1}^{n} \lambda_j x_j$. We know that all the $\lambda_j \geq 0$ and $\sum \lambda_j = 1$. Define random variable

$$X = x_j \text{ with probability } \lambda_j.$$

Clearly, $\mathbb{E}[X] = z$. Let P be the probability distribution of X. Take $X_1, \dots X_k$ to be i.i.d samples from P. Then, get the average $\bar{X}_k$. We have

$$\mathbb{E}[|\bar{X}_k - z|^2] = \frac{1}{k}\mathbb{E}[|X - \mathbb{E}[X]|_2^2] \leq \frac{1}{k}\mathbb{E}[|X|_2^2] \leq \frac{1}{k}.$$

Thus, there has to be positive mass between 0 and $\frac{1}{k}$, which means that there exists a choice of samples that realizes this bound and we are done.

§2.1 Application to Polytope

Proposition 2.5

Let P be a polytope with n vertices and diameter ≤ 2. Then, the covering number N_ϵ of P w.r.t L_2 norm satisfies

$$\log N_\epsilon \leq \frac{c \log n}{\epsilon^2}.$$

Proof. Let $x_1, \dots x_n$ denote vertices of P. Center P so that $P \subset B_2$. Take

$$N = \left\{ \frac{1}{k} \sum_{j=1}^{k} z_j \mid z_j \in \{x_1, \dots x_n\} \right\}.$$

Note that N is a $1/\sqrt{k}$ net for P. Take $k = \frac{1}{\epsilon^2}$. Then, we get that N is an ϵ-net with size n^k. Thus,

$$\log |N| = k \log n = \frac{\log n}{\epsilon^2}.$$

§2.2 Frank-Wolfe: An Algorithmic Approach

This section looks at how we can actually compute the set of vectors and weights that approximates z.

The idea here is that if $z \in C$, then $\min_{w \in C} |z - w|^2 = 0$.

Optimization Algorithm: Iterate to get $w_1, \dots w_k$ s.t. $|z - w_k|^2 \leq \frac{c}{k}$.

Assume WLOG that 0 is a vertex of C.

Theorem 2.6

(Frank-Wolfe Algorithm)

Initialize $w_0 = 0$.

Iterate $y_k = \arg\min_{x \in C} \langle x - w_k, \nabla f(w_k) \rangle$.

$$w_{k+1} = w_k + \frac{2}{2+k}(y_k - w_k)$$

We can verify that each w_k is in the convex hull of the S vertices.

Proof. We will show that for f convex, if $f(w_k) - \min_{w \in C} f(w) \leq \frac{c}{k}$. Assume that $\min_{w \in C} f(w) = f(w^*) = 0$. Then,

$$\begin{aligned} f(w_{k+1}) &= f(w_k + \frac{2}{2+k}(y_k - w_k)) \\ &\leq f(w_k) + \frac{2}{2+k}(y_k - w_k)^T \nabla f(w_k) + \frac{c}{k^2} \end{aligned}$$

We know that $(y_k - w_k)^T \nabla f(w_k) < (x - w_k)^T \nabla f(w_k)$ for all x, so by convexity,

$$(y_k - w_k)^T \nabla f(w_k) < (x - w_k)^T \nabla f(w_k) \leq f(x) - f(w_k).$$

Take $x = w^*$ such that $f(x) = f(w^*) = 0$. Then, we have

$$f(w_{k+1}) \leq f(w_k) - \frac{2}{2+k} f(w_k) + \frac{c}{k^2} = \frac{k}{2+k} f(w_k) + \frac{c}{k^2}$$

and then we can just conclude by induction.