# IEOR 265 – Lecture 2
# Complexity Measures

---

# 1 Metric Entropy

In the previous lecture, we discussed how the majority of volume in a high-dimensional convex body is concentrated in a small radius. However, this does not completely characterize the important properties of convex bodies. For instance, consider the unit $\ell_1$- and $\ell_\infty$-balls, that is

$$\mathcal{B}_1 = \{x : \|x\|_1 \leq 1\}$$
$$\mathcal{B}_\infty = \{x : \|x\|_\infty \leq 1\},$$

where $\|x\|_1 = \sum_{j=1}^p |x_i|$ is the $\ell_1$-norm and $\|x\|_\infty = \max_j |x_i|$ is the $\ell_\infty$-norm. Though these balls have the same unit radius, the $\ell_1$-ball $\mathcal{B}_1$ has $2p$ vertices whereas the $\ell_\infty$-ball $\mathcal{B}_\infty$ has $2^p$ vertices. In other words, the polytope $\mathcal{B}_\infty$ is significantly more complex than $\mathcal{B}_1$.

Given this large discrepancy in the complexity of these two balls, it is natural to ask whether it is possible to define some measure of complexity for general convex bodies? In fact, there is a broader question of whether it is possible to define some measure of complexity for general subsets of $\mathbb{R}^p$? We cannot simply count the number of vertices of the subsets, because in general these sets will not be polytopes. It turns out that the answer to the above questions is an emphatic "yes". The situation is in fact more complex, because it turns out that there are several interesting notions of complexity for subsets of $\mathbb{R}^p$.

## 1.1 Definition of Metric Entropy

One simple notion of complexity is known as the "covering number" or "metric entropy" (they are related by a logarithm). Given a set $\mathcal{T} \subset \mathbb{R}^p$ and $\mathcal{L} \subset \mathbb{R}^p$, the covering number $N(\mathcal{T}, \mathcal{L})$ is the minimum number of translates of $\mathcal{L}$ needed to cover $\mathcal{T}$. Then, the metric entropy is defined as $\log N(\mathcal{T}, \mathcal{L})$. An important special case is when $\mathcal{L} = \epsilon\mathcal{B}_2$, where $\mathcal{B}_2 = \{x : \|x\|_2 \leq 1\}$ is the unit $\ell_2$-ball. Unfortunately, it is difficult to get accurate bounds on the covering number or metric entropy for a general set $\mathcal{T}$.

## 1.2 Covering Number Estimate

One approach to bounding the covering number is to compare the volumes of the sets $\mathcal{T}, \mathcal{L}$. In particular, we have that

$$\frac{\mathsf{vol}(\mathcal{T})}{\mathsf{vol}(\mathcal{L})} \leq N(\mathcal{T}, \mathcal{L}) \leq \frac{\mathsf{vol}(\mathcal{T} \oplus \frac{1}{2}\mathcal{L})}{\mathsf{vol}(\frac{1}{2}\mathcal{L})},$$

where $\mathcal{U} \oplus \mathcal{V} = \{u + v : u \in \mathcal{U}, v \in \mathcal{V}\}$ is the Minkowski sum. A useful corollary of this result is that if $\mathcal{T}$ is a symmetric convex set, then for every $\epsilon > 0$ we have

$$\Big(\frac{1}{\epsilon}\Big)^p \leq N(\mathcal{T}, \epsilon\mathcal{T}) \leq \Big(2 + \frac{1}{\epsilon}\Big)^p.$$

### 1.3 Example: Covering Number of $\ell_2$-Ball

Consider the unit $\ell_2$-ball $\mathcal{B}_2 = \{x : \|x\|_2 \leq 1\}$. Then the covering number of $\mathcal{B}_2$ by $\epsilon_2$ is bounded by

$$\Big(\frac{1}{\epsilon}\Big)^p \leq N(\mathcal{B}_2, \epsilon\mathcal{B}_2) \leq \Big(2 + \frac{1}{\epsilon}\Big)^p.$$

For the $\ell_2$-ball with radius $\lambda$, which is defined as $\mathcal{B}_2(\lambda) = \{x : \|x\|_2 \leq \lambda\}$, we thus have that the covering number of $\mathcal{B}_2(\lambda)$ by $\epsilon\mathcal{B}_2$ is bounded by

$$\lambda^p \cdot \Big(\frac{1}{\epsilon}\Big)^p \leq N(\mathcal{B}_2, \epsilon\mathcal{B}_2) \leq \lambda^p \cdot \Big(2 + \frac{1}{\epsilon}\Big)^p.$$

This is exponential in $p$.

## 2 Gaussian Average

Another interesting notion of complexity is known as a "Gaussian average" or "Gaussian width". Given a set $\mathcal{T} \subset \mathbb{R}^p$, we define the Gaussian average as

$$w(\mathcal{T}) = \mathbb{E}\Big(\sup_{t \in \mathcal{T}} g' t\Big),$$

where $g \in \mathbb{R}^p$ is a Gaussian random vector with zero-mean and identity covariance matrix (or equivalently a random vector where each entry is an iid Gaussian random variable with zero-mean and unit variance). Unfortunately, computing the Gaussian average for a given set $\mathcal{T}$ can be difficult unless $\mathcal{T}$ has some simple structure.

### 2.1 Invariance Under Convexification

One of the most important properties (from the standpoint of high-dimensional statistics) of the Gaussian average is that

$$w(\text{conv}(\mathcal{T})) = w(\mathcal{T}),$$

where conv$(\mathcal{T})$ denotes the convex hull of $\mathcal{T}$. The proof of this is quite simple and instructive: First, note that $w(\text{conv}(\mathcal{T})) \geq w(\mathcal{T})$, since $\mathcal{T} \subseteq \text{conv}(\mathcal{T})$. Second, note that if $t \in \text{conv}(\mathcal{T})$ then by Carathéodory's theorem it can be represented as $t = \sum_{j=1}^{p+1} \mu_j t_j$ where $\mu_j \in [0,1]$, $\sum_j \mu_j = 1$, and $t_j \in \mathcal{T}$. As a result, we have

$$\begin{aligned} w(\mathsf{conv}(\mathcal{T})) &= \mathbb{E}\Big( \sup_{\mu_j \in [0,1], \sum_j \mu_j = 1, t_j \in \mathcal{T}} g'(\textstyle\sum_{j=1}^{p+1} \mu_j t_j)\Big) \\ &\leq \mathbb{E}\Big( \sup_{\mu_j \in [0,1], \sum_j \mu_j = 1, t_j \in \mathcal{T}} \max_j g' t_j \Big) = \mathbb{E}\Big( \sup_{t_j \in \mathcal{T}} \max_j g' t_j \Big) \\ &= \mathbb{E}\Big( \sup_{t \in \mathcal{T}} g' t \Big) = w(\mathcal{T}). \end{aligned}$$

Since we have $w(\mathsf{conv}(\mathcal{T})) \geq w(\mathcal{T})$ and $w(\mathsf{conv}(\mathcal{T})) \leq w(\mathcal{T})$, it must be the case that $w(\mathsf{conv}(\mathcal{T})) = w(\mathcal{T})$. Note that this equivalence also follows by noting that the maximum of a convex function over a closed convex set is attained at an extreme point of the convex set.

## 2.2 Sudakov's Minoration

It turns out there is a relationship between the Gaussian average and the metric entropy of a set. One such relationship is known as Sudakov's minoration. In particular, if $\mathcal{T}$ is a symmetric set (i.e., if $t \in \mathcal{T}$, then $-t \in \mathcal{T}$), then we have

$$\sqrt{\log N(\mathcal{T}, \epsilon \mathcal{B}_2)} \leq C \cdot \frac{w(\mathcal{T})}{\epsilon},$$

where $C$ is an absolute constant. This is a useful relationship because it allows us to upper bound the metric entropy of a set if we can compute its Gaussian average.

## 2.3 Example: Gaussian Average of $\ell_2$-Balls

Consider the set $\mathcal{B}_2(\lambda) = \{x : \|x\|_2 \leq \lambda\}$. Its Gaussian average is defined as

$$w(\mathcal{B}_2(\lambda)) = \mathbb{E}\Big( \sup_{t \in \mathcal{B}_2(\lambda)} g' t \Big).$$

The quantity $g't$ is maximized whenever $t$ is in the direction of $g$ (i.e., $t \sim \frac{g}{\|g\|_2}$) and has length $\|t\|_2 = \lambda$. Thus, the quantity is maximized for $t = \lambda \cdot \frac{g}{\|g\|_2}$. As a result, the Gaussian average is

$$w(\mathcal{B}_2(\lambda)) = \lambda \cdot \mathbb{E}(\|g\|_2).$$

Since $\|g\|_2$ has a chi-distribution, standard results about this distribution give that

$$c\lambda\sqrt{p} \leq w(\mathcal{B}_2(\lambda)) \leq \lambda\sqrt{p},$$

where $c$ is an absolute positive constant. Finally, we can use Sudakov's minoration to upper bound the covering number:

$$N(\mathcal{B}_2(\lambda), \epsilon \mathcal{B}_2) \leq \exp\Big(\frac{C^2 p}{\epsilon^2}\Big).$$

This is exponential in $p$, just like the previous bound.

## 2.4 Example: Gaussian Average of $\ell_1$-Balls

Consider the set $\mathcal{B}_1(\lambda) = \{x : \|x\|_1 \le \lambda\}$. Its Gaussian average is defined as

$$w(\mathcal{B}_1(\lambda)) = \mathbb{E}\Big(\sup_{t\in\mathcal{B}_1(\lambda)} g't\Big).$$

By Hölder's inequality, we know that $g't \le |g't| \le \|g\|_\infty \cdot \|t\|_1$. Thus, we can bound the Gaussian average by

$$w(\mathcal{B}_1(\lambda)) = \mathbb{E}\Big(\sup_{t\in\mathcal{B}_1(\lambda)} g't\Big) \le \mathbb{E}\Big(\sup_{t\in\mathcal{B}_1(\lambda)} \|g\|_\infty \cdot \|t\|_1\Big) = \lambda \cdot \mathbb{E}\Big(\max_j |g_j|\Big).$$

Hence, the question is how to upper bound $\mathbb{E}\Big(\max_j |g_j|\Big)$.

1. The first observation is that

$$\max_j |g_j| = \max\Big(\max_j g_j, \max_j -g_j\Big).$$

2. The second observation is that if $V_i$ for $i = 1, \dots, n$ are (not necessarily independent) Gaussian random variables with zero-mean and variance $\sigma^2$, then we have

$$\exp(u\mathbb{E}(\max_i V_i)) \le \mathbb{E}(\exp(u \cdot \max_i V_i)) = \\ \mathbb{E}\max_i \exp(uV_i) \le \sum_{i=1}^n \mathbb{E}(\exp(uV_i)) \le n\exp(u^2\sigma^2/2),$$

where the first inequality follows from Jensen's inequality and the last equality is an identity from the definition of the moment generating function of a Gaussian. Rearranging terms gives

$$\mathbb{E}(\max_i V_i) \le \frac{\log n}{u} + \frac{u\sigma^2}{2}.$$

We can tighten this bound by choosing $u$ to minimize the right hand side of the above equation. The minimizing value makes the derivative equal to zero, meaning

$$\frac{-\log n}{u^2} + \frac{\sigma^2}{2} = 0 \Rightarrow u = \sqrt{2\log n}/\sigma.$$

Substituting this value of $u$ into the upper bound yields

$$\mathbb{E}(\max_i V_i) \le \sigma\sqrt{2\log n}.$$

Since $\sigma^2 = 1$ for $g_j$ and $-g_j$, combining these two observations gives that

$$\mathbb{E}\Big(\max_j |g_j|\Big) \le \sqrt{2\log 2p} \le \sqrt{2\log 2 + 2\log p} \le \sqrt{4\log p},$$

whenever $p \geq 2$.

Consequently, we have that the Gaussian average of the $\ell_1$-ball is

$$w(\mathcal{B}_1(\lambda)) \leq \lambda \cdot \sqrt{4 \log p}.$$

whenever $p \geq 2$. We can use Sudakov's minoration to upper bound the covering number:

$$N(\mathcal{B}_1(\lambda), \epsilon\mathcal{B}_2) \leq \exp\Big(\frac{C^2\lambda^2 \cdot 4\log p}{\epsilon^2}\Big) = (\exp(\log p))^{4C^2\lambda^2/\epsilon^2} = p^{c\lambda^2/\epsilon^2},$$

where $c$ is an absolute constant. Interestingly, this is polynomial in $p$, which is in contrast to the covering number of $\mathcal{B}_2(\lambda)$.

## 2.5 Example: Gaussian Average of $\ell_\infty$-Balls

Consider the set $\mathcal{B}_\infty(\lambda) = \{x : \|x\|_\infty \leq \lambda\}$. Its Gaussian average is defined as

$$w(\mathcal{B}_\infty(\lambda)) = \mathbb{E}\Big(\sup_{t \in \mathcal{B}_\infty(\lambda)} g't\Big).$$

By Hölder's inequality, we know that $g't \leq |g't| \leq \|g\|_1 \cdot \|t\|_\infty$. Thus, we can bound the Gaussian average by

$$w(\mathcal{B}_\infty(\lambda)) = \mathbb{E}\Big(\sup_{t \in \mathcal{B}_\infty(\lambda)} g't\Big) \leq \mathbb{E}\Big(\sup_{t \in \mathcal{B}_\infty(\lambda)} \|g\|_1 \cdot \|t\|_\infty\Big) = \lambda \cdot \mathbb{E}\Big(\|g\|_1\Big).$$

But note that $\|g\|_1 = \max_{s_i \in \{-1,1\}} \sum_{i=1}^p s_i g_i$. Since the $g_i$ are iid Gaussians with zero-mean and unit variance, for fixed $s_i$ the quantity $\sum_{i=1}^p s_i g_i$ is a Gaussian with zero mean and variance $p$. Thus, using our earlier bound gives

$$\mathbb{E}(\|g\|_1) = \mathbb{E}(\max_{s_i \in \{-1,1\}} \sum_{i=1}^p s_i g_i) \leq \sqrt{4p \log 2^p} = \sqrt{4 \log 2 \cdot p^2}.$$

Thus, the Gaussian average is

$$w(\mathcal{B}_\infty(\lambda)) \leq \sqrt{4 \log 2}\lambda \cdot p.$$

We can use Sudakov's minoration to upper bound the covering number:

$$N(\mathcal{B}_\infty(\lambda), \epsilon\mathcal{B}_2) \leq \exp\Big(\frac{C^2\lambda^2 \cdot 4\log 2 \cdot p^2}{\epsilon^2}\Big),$$

where $c$ is an absolute constant. This is exponential in $p$.

# 3 Rademacher Average

Another notion of complexity is known as a "Rademacher average" or "Rademacher width". Given a set $\mathcal{T} \subset \mathbb{R}^p$, we define the Rademacher average as

$$r(\mathcal{T}) = \mathbb{E}\Big(\sup_{t\in\mathcal{T}} \epsilon' t\Big),$$

where $\epsilon \in \mathbb{R}^p$ is a Rademacher random vector, meaning each component $\epsilon_i$ is independent and

$$\mathbb{P}(\epsilon_i = \pm 1) = \frac{1}{2}.$$

Similar to the Gaussian average, computing the Rademacher average for a given set $\mathcal{T}$ can be difficult unless $\mathcal{T}$ has some simple structure. Though many results for Gaussian averages have analogs for Rademacher averages, Rademacher averages are more difficult to work with because many comparison results for Gaussian averages do not have analogs for Rademacher averages.

Furthermore, it is the case that Rademacher and Gaussian averages are equivalent:

$$c \cdot r(\mathcal{T}) \le w(\mathcal{T}) \le C \cdot \log p r(\mathcal{T}),$$

where $c, C$ are absolute constants.

# 4 $M^*$ Bound

The reason we are interested in complexity measures of sets is because of the following result, which is known as the $M^*$ Bound. If (i) $\mathcal{T} \subset \mathbb{R}^p$ is bounded and symmetric (i.e., $\mathcal{T} = -\mathcal{T}$), and (ii) $\mathcal{E}$ is a random subspace of $\mathbb{R}^p$ with fixed codimension $n$ and drawn according to an appropriate distribution, then

$$\mathbb{E}\Big(\text{diam}(\mathcal{T} \cap \mathcal{E})\Big) \le \frac{Cw(\mathcal{T})}{\sqrt{n}}.$$

.

# 5 More Reading

The material in these sections follows that of


R. Vershynin (2014) *Estimation in high dimensions: a geometric perspective*, in Sampling Theory, a Renaissance. Springer. To appear.

R. Vershynin (2009) *Lectures in geometric functional analysis*. Available online: http://www-personal.umich.edu/~romanv/papers/GFA-book/GFA-book.pdf.

M. Ledoux and M. Talagrand (1991) *Probability in Banach Spaces*. Springer.

More details about these sections can be found in the above references.