# Math 254A Lecture 8 Notes

Daniel Raban

April 14, 2021

## 1 Integral Formula for the Fenchel-Legendre Transform

### 1.1 The Fenchel-Legendre transform and the integral formula

Last time, we defined the **Fenchel-Legendre transform** $s^* = \sup_x s(x) + \langle y, x \rangle$, which is convex, lower semicontinuous, is $s^* : X^* \to (-\infty, \infty]$, and is not always $+\infty$.[1] We also saw that $s = (s^*)_*$, so we can recover $s$ from its Fenchel-Legendre transform.

Let's focus on the $X = Y^*$ case, since this also subsumes the $X = \mathbb{R}^k$ case. Also assume $\lambda \neq 0$.

**Theorem 1.1.** *In this generalized type counting problem for $X = Y^*$,*

$$s^*(y) = \log \int e^{\langle y, \varphi \rangle} \, d\lambda$$

*for $y \in Y$.*

Before proving this, observe:

$$\begin{aligned} \exp(s^*(ty + (1-t)w)) &= \int e^{t\langle y, \varphi \rangle + (1-t)\langle w, \varphi \rangle} \, d\lambda \\ &= \int e^{t\langle y, \varphi \rangle} \cdot e^{(1-t)\langle w, \varphi \rangle} \, d\lambda \end{aligned}$$

Using Hölder's inequality,

$$\leq \left( \int e^{\langle y, \varphi \rangle} \, d\lambda \right)^t \left( \int e^{\langle w, \varphi \rangle} \, d\lambda \right)^{1-t},$$

so taking logs gives that this expression is convex. We can also check that this expression is lower semicontinuous.

[1] Many authors study $\widetilde{s} = -s$ throughout and then get $s^*(y) = \sup_x \langle y, x \rangle - \widetilde{s}(x)$ and $\widetilde{s}(z) = \sup_y \langle y, z \rangle - s^*(y)$. We use a different convention.

## 1.2 Proofs of the upper bound and the lower bound

*Proof.* ($\leq$): Since $s^*(y) = \sup_x s(x) + \langle y, x\rangle$, we need to show that

$$s(x) + \langle y, x\rangle \leq \log \int e^{\langle y,\varphi\rangle}\, d\lambda$$

for all $x$. Let $\varepsilon > 0$, and consider $U = \{x' : \langle y, x'\rangle > \langle y, x\rangle - \varepsilon\}$. We know that

$$\begin{aligned}
s^{n\cdot(s(x)+\langle y,x\rangle)} &\leq e^{n(s(U)+\langle y,x\rangle)} \\
&= e^{o(n)} e^{n\langle y,x\rangle} \lambda^{\times n}\left(\left\{p : \frac{1}{n}\sum_{i=1}^n \varphi(p_i) \in U\right\}\right) \\
&= e^{o(n)} e^{n\langle y,x\rangle} \lambda^{\times n}\left(\left\{p : \sum_{i=1}^n \langle y, \varphi(p_i)\rangle > n\langle y, x\rangle - n\varepsilon\right\}\right)
\end{aligned}$$

Exponentiate both sides in the inequality and apply Markov's inequality:[2]

$$\begin{aligned}
&\leq e^{o(n)} e^{n\langle y,x\rangle} e^{n\varepsilon - n\langle y,x\rangle} \int e^{\sum_{i=1}^n \langle y,\varphi(p_i)\rangle}\, d\lambda^{\times n} \\
&= e^{o(n)+n\varepsilon} \int_{M^n} \prod_{i=1}^n e^{\langle y,\varphi(p_i)\rangle}\, d\lambda^n \\
&= e^{o(n)} e^{\varepsilon n}\left(\int e^{\langle y,\varphi\rangle}\, d\lambda\right)^n,
\end{aligned}$$

so

$$n(s(U) + \langle y, x\rangle) \leq o(n) + \varepsilon n + n \log \int e^{\langle y,\varphi\rangle}\, d\lambda.$$

Divide by $n$ and send $n \to \infty$ to get

$$s(x) + \langle y, x\rangle \leq \ s(U) + \langle y, x\rangle \leq \varepsilon + \log \int e^{\langle y,\varphi\rangle}\, d\lambda.$$

Since $\varepsilon$ is arbitrary, we get ($\leq$).

To get the lower bound, let's look at the proof of the upper bound and try to make it as tight as possible. The first inequality is close if $U$ is a small enough neighborhood of $x$. In the Chernoff bound, we want to see when this is close to equality. To prove ($\geq$), we will look at the Chernoff bound step; here's the idea: Consider

$$e^{n\langle y,x\rangle} \lambda^{\times n}\left(\left\{p : \frac{1}{n}\sum_{i=1}^n \langle y, \varphi(p_i)\rangle \in U\right\}\right),$$

[2] This is sometimes called a **Chernoff bound**.

where we want to make $U$ small enough around $x$ to force this to be $\approx \langle y, x\rangle$. We then get

$$e^{n\langle y,x\rangle}\lambda^{\times n}\left(\left\{p : \exp\sum_{i=1}^{n}\langle y, \varphi(p_i)\rangle \approx e^{n\langle y,x\rangle}, \frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \in U\right\}\right).$$

This is

$$\approx e^{\pm\varepsilon n}\int_{\{\frac{1}{n}\sum_{i=1}^{n}\varphi(p_i)\in U\}} e^{\sum_{i=1}^{n}\langle y,\varphi(p_i)\rangle}\,d\lambda^{\times n} \le e^{\varepsilon n}\int_{M^n} e^{\sum_{i=1}^{n}\langle y,\varphi(p_i)\rangle}\,d\lambda^{\times n}.$$

So the question becomes: Can we find an $x$ where most of the mass lies in the set $\{\frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \in U\}$?

Now let's prove $(\geq)$ carefully. First assume two conditions:

1. $Z = \int e^{\langle y,\varphi\rangle}\,d\lambda < \infty$.

2. $p$ takes values in a compact subset $K$ of $X$.

In this case, we can define a new probability measure on $M$ by

$$d\theta(p) = \frac{1}{Z}e^{\langle y,\varphi(p)\rangle}\,d\lambda(p)$$

(using assumption 1). Now, for any $A \subseteq M^n$,

$$\int_A e^{\sum_{i=1}^{n}\langle y,\varphi(p_i)\rangle}\,d\lambda^{\times n} = Z^n\theta^{\times n}(A).$$

With $A = \{\frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \in U\}$, we get

$$Z^n\theta^{\times n}\left(\left\{\frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \in U\right\}\right)$$

This suggests we can use the Weak Law of Large Numbers for $\theta$ and $\varphi$.[3] To do this carefully, we need assumption 2: $p$ takes values in $K \subseteq X$, so it has a barycenter with respect to $\theta$: a unique $x \in K$ such that

$$\int \langle y, \varphi\rangle\,d\theta = \langle y, x\rangle \qquad \forall y \in Y.$$

And now a vector-valued Weak Law of Large Numbers holds: for this $x$ and any weak* neighborhood $U \ni x$, we get

$$\theta^{\times n}\left(\left\{\frac{1}{n}\sum_{i=1}^{n}\varphi(p_i) \in U\right\}\right) = 1 - o(1)$$

[3] This is the key idea of the lower bound proof. It is called the **change of measure** idea.

as $n \to \infty$. As a result, for any weak* neighborhood of this $x$, we now get

$$\int_{\{\frac{1}{n}\sum_{i=1}^n \varphi(p_i) \in U\}} = Z^n \theta^{\times n}\left(\left\{\frac{1}{n}\sum_{i=1}^{n} \varphi(p_i) \in U\right\}\right) \geq Z^n e^{o(n)}.$$

Insert this to reverse the previous upper bound proof to get an $x$ such that $s(x) + \langle y, x\rangle \geq \log Z - \varepsilon$. This gives $s^*(y) = \log Z$.

To remove assumptions 1 and 2, recall that $(M, \lambda)$ is $\sigma$-finite and $X = \bigcup_n K_n$, so for any $a < \int e^{\langle y,\varphi\rangle}\, d\lambda$, there exists a measurable $A \subseteq M$ such that $\infty > \int_A e^{\langle y,\varphi\rangle}\, d\lambda > a$, and $\varphi(A)$ takes values in some $K_n$. Now run the previous argument with $d\lambda'(p) = \mathbb{1}_A(p)\, d\lambda(p)$ to get that for every $\varepsilon$, there is an $x$ such that $s(x) + \langle y, x\rangle \geq \log a - \varepsilon$. Since $a < \int e^{\langle y,\varphi\rangle}\, d\lambda$ was arbitrary, we get

$$s^*(y) = \log \int e^{\langle y,\varphi\rangle}\, d\lambda,$$

even if this is $+\infty$. $\square$