PHY 521: Problem Set 1

Due on February 16, 2023

Professor Michael Aizenman

Evan Dogariu

Collaborators: Beyer White, Ben Babel

Problem 2.1

Show that the Legendre transform of the entropy function which was computed in (1.6) (using Stirling's approximation) is consistent with (2.11).

Solution

Proof. We have that

$$s(n) = -\left(n\ln(n) + (1-n)\ln(1-n)\right)$$

and we wish to show that

$$\sup_n[\widetilde{\mu}\cdot n + s(n)] = \ln\left(1 + e^{\widetilde{\mu}}\right)$$

Note first that since $s(n)$ is bounded and strictly concave, $\widetilde{\mu}\cdot n + s(n)$ is also bounded and strictly concave when viewed as a function of $n \in [0,1]$. Therefore, there is exactly one critical point, corresponding to the maximal value (the endpoints $n \in \{0,1\}$ are certainly not maximal, as $s(0) = s(1) = 0$). So, we seek the unique solution to

$$\frac{d}{dn}\left[\widetilde{\mu}\cdot n + s(n)\right] = 0,$$

which will be the value of n at which the supremum is achieved. We calculate

$$\begin{aligned}
\frac{d}{dn}\left[\widetilde{\mu}\cdot n + s(n)\right] &= \widetilde{\mu} + s'(n) = \widetilde{\mu} - \left(\frac{n}{n} + \ln(n) - \frac{1-n}{1-n} - \ln(1-n)\right) \\
&= \widetilde{\mu} - \left(1 - 1 + \ln\left(\frac{n}{1-n}\right)\right) = \widetilde{\mu} - \ln\left(\frac{n}{1-n}\right)
\end{aligned}$$

Setting this to 0, we find

$$\frac{n}{1-n} = e^{\widetilde{\mu}} \implies n = (1-n)e^{\widetilde{\mu}} \implies n = \frac{e^{\widetilde{\mu}}}{1+e^{\widetilde{\mu}}}$$

This is the value of n at which the supremum is achieved. To find the actual value of the supremum, we plug this in to see that

$$\begin{aligned}
\sup_n[\widetilde{\mu}\cdot n + s(n)] &= \widetilde{\mu}\cdot\frac{e^{\widetilde{\mu}}}{1+e^{\widetilde{\mu}}} - \left(\frac{e^{\widetilde{\mu}}}{1+e^{\widetilde{\mu}}}\ln\left(\frac{e^{\widetilde{\mu}}}{1+e^{\widetilde{\mu}}}\right) + \frac{1}{1+e^{\widetilde{\mu}}}\ln\left(\frac{1}{1+e^{\widetilde{\mu}}}\right)\right) \\
&= \frac{1}{1+e^{\widetilde{\mu}}}\left(\widetilde{\mu}e^{\widetilde{\mu}} - e^{\widetilde{\mu}}\ln\left(e^{\widetilde{\mu}}\right) + e^{\widetilde{\mu}}\ln\left(1+e^{\widetilde{\mu}}\right) + \ln\left(1+e^{\widetilde{\mu}}\right)\right) \\
&= \frac{1}{1+e^{\widetilde{\mu}}}\left(1+e^{\widetilde{\mu}}\right)\ln\left(1+e^{\widetilde{\mu}}\right) = \ln\left(1+e^{\widetilde{\mu}}\right)
\end{aligned}$$

as desired. ■

Problem 3.1

Prove the following theorem:

Theorem 1. *Let F_n be a sequence of convex functions over a common open interval $I \to \mathbb{R}$ which are differentiable and converge pointwise as $n \to \infty$. Then for every $x \in I$ at which $F(x) = \lim_{n\to\infty} F_n(x)$ is differentiable,*

$$F'(x) = \lim_{n\to\infty} F_n'(x)$$

Furthermore, regardless of the existence of $F'(x)$ the following relation holds

$$F'_-(x) \leq \liminf_{n\to\infty} F_n'(x) \leq \limsup_{n\to\infty} F_n'(x) \leq F'_+(x)$$

Solution

Proof. Fix $x \in I$. Fix $h > 0$ (make sure that h is small enough that both $x - h, x + h \in I$). We know from the notes that convexity is equivalent to monotonicity of the chord slopes. Thus, the convexity of each F_n easily yields that for every $n \in \mathbb{N}$,

$$\frac{F_n(x) - F_n(x-h)}{h} \leq F_n'(x) \leq \frac{F_n(x+h) - F_n(x)}{h}$$

Let us focus in on the right inequality $F_n'(x) \leq \frac{F_n(x+h)-F_n(x)}{h}$. Since this holds for every n, it will also hold under performing a limsup. So, taking the limsup as $n \to \infty$ (and similarly taking the liminf on the left inequality), we get that

$$\liminf_{n\to\infty} \frac{F_n(x) - F_n(x-h)}{h} \leq \liminf_{n\to\infty} F_n'(x) \leq \limsup_{n\to\infty} F_n'(x) \leq \limsup_{n\to\infty} \frac{F_n(x+h) - F_n(x)}{h},$$

where the middle inequality holds trivially. However, the pointwise convergence $F_n(x) \to F(x)$ at x guarantees that the liminf and limsup of the chord slopes converge to the appropriate pointwise limit. In other words,

$$\frac{F(x) - F(x-h)}{h} \leq \liminf_{n\to\infty} F_n'(x) \leq \limsup_{n\to\infty} F_n'(x) \leq \frac{F(x+h) - F(x)}{h}$$

Note that this result holds for any arbitrary $h > 0$. Therefore, it also holds in the limit as $h \to 0$, yielding that for our fixed x,

$$F'_-(x) \leq \liminf_{n\to\infty} F_n'(x) \leq \limsup_{n\to\infty} F_n'(x) \leq F'_+(x)$$

This holds for an arbitrary $x \in I$, thus proving the second relation of the theorem.

Now, at every point $x \in I$ for which F is differentiable, we have that $F'_-(x) = F'_+(x)$. Therefore, the above inequalities must be equality, guaranteeing that for such x,

$$F'_-(x) = \liminf_{n\to\infty} F_n'(x) = \limsup_{n\to\infty} F_n'(x) = F'_+(x)$$

So, the limit $\lim_{n\to\infty} F_n'(x)$ exists, and must equal $F'(x)$ for such x, as desired. ■

Problem 3.2

Sketch a function $G : \mathbb{R} \to \mathbb{R}$ with a discontinuous derivative. How is this singularity expressed in the function's Legendre transform? Sketch, and prove.

Solution

Proof. Consider the absolute value function $G(x) = |x|$. We can explicitly write the Legendre transform TG. Seeking to describe

$$(TG)(y) = \sup_x [y \cdot x - G(x)] = \sup_x [y \cdot x - |x|],$$

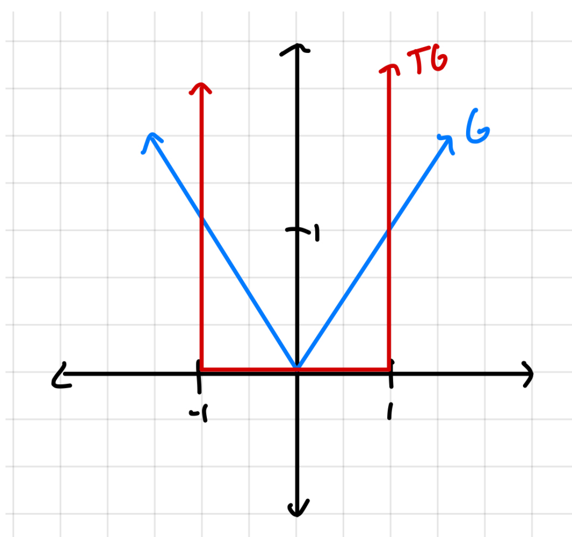
we note that for $y = 0$, the supremum is clearly achieved at $x = 0 \implies (TG)(0) = 0$, since $|x|$ is always nonnegative. Next, for $y \in [-1, 1]$, we similarly achieve the supremum at $x = 0 \implies (TG)(y) = 0$ for $y \in [-1, 1]$. To see this, note that if $|y| \leq 1$, we have that

$$yx - |x| \leq 0$$

for all x, and the value of 0 is achieved at $x = 0$. Lastly, note that if $y > 1$, we can always select a more positive x to increase the value of $yx - |x|$, and so the supremum is $+\infty$. Similarly, if $y < -1$, we can always select a more negative x to increase the value of $yx - |x|$. This fully determines the function TG to be

$$(TG)(y) = \begin{cases} 0 & |y| \leq 1 \\ +\infty & else \end{cases}$$

Sketches of G and TG are given below in blue and red, respectively.



In this example, we observe that the function G has a discontinuous derivative: to the left of $x = 0$ the derivative is -1, and to the right it is +1. This is handled exactly how we would expect it to be in the Legendre transform TG, where there is a flat linear section connecting the two discontinuous values (± 1). In general, we know that a kink in G (discontinuous derivative at a point) corresponds to a flat linear section of TG between the values of the derivative before and after the discontinuity, exactly as we see here! ∎

Problem 3.3

Describe the Legendre transform of a function which fails to be convex in a strict subset of its domain of definition, as depicted in Fig. 3.4.

Solution

Proof. Consider a setting such as that of Fig. 3.4, which is shown below for convenience.

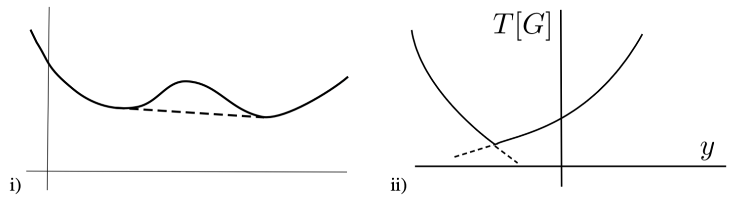


Figure 3.4 i) the "convex hull" of a non-convex function, and ii) a sketch of the resulting function's Legendre transform.

Let (x_1, x_2) be the interval over which the function G fails to be convex. We start by noting that any function that agrees with G outside this interval and also lies above the dotted line on this interval will have the same Legendre transform. This can be seen either through the fact that the Legendre transform is convex and the use of the Legendre transform's involutive property, or by reasoning about sweeping y. We note that the Legendre transform

$$(TG)(y) = \sup_x [yx - G(x)]$$

will find the largest value of x that has a line of support with slope y. For all such functions described above, the Legendre transform will select points naturally until we get to the value $y = G'(x_1) - \epsilon$. Once we cross $y = G'(x_1)$, the supremum will jump to selecting x_2, and continue from there. In terms of the value of the Legendre transform, this yields a continuous curve (as the values of the supremum match at x_1 and x_2 since they have the same line of support), but with a discontinuous derivative. More precisely, we find a kink in the graph of TG at a value $y_0 = \frac{G(x_2)-G(x_1)}{x_2-x_1}$ (the slope of the supporting line of both x_1 and x_2), where the left derivative equals x_1 and the right derivative equals x_2. Applying the Legendre transform to this graph TG will yield a convex curve with a straight line between $x_1 = (TG)'_-(y_0)$ and $x_2 = (TG)'_+(y_0)$ - this curve is precisely the convex hull of any of the functions described earlier.

To sum up, the Legendre transform TG of a function G failing to be convex in an interval (x_1, x_2) will be continuous but have a kink at $y = \frac{G(x_2)-G(x_1)}{x_2-x_1}$, where the left derivative of TG equals x_1 and the right derivative equals x_2. Taking the Legendre transform of TG again yields the convex hull of our original curve.

■

Problem 3.4

Prove the following theorem:

Theorem 2. *For any strictly convex, differentiable function $G : \mathbb{R}^\nu \to \mathbb{R} \cup \{+\infty\}$ for which TG is also strictly convex and differentiable, we have that*

$$T[TG] = G$$

Solution

Proof. Let $E = \{\vec{y} \in \mathbb{R}^\nu \ : \ \vec{\nabla} G(\vec{x}) = \vec{y} \text{ for some } \vec{x} \in \mathbb{R}^\nu\}$ be the set of all gradients attained by G. Define $X : E \to \mathbb{R}^\nu$ to be the map such that for all $\vec{y} \in E$, we have $\vec{\nabla} G(X(\vec{y})) = \vec{y}$. In other words, X maps gradients $\vec{y}$ that are achieved by G to the points $\vec{x}$ at which G achieves them. Note that strict convexity of G implies that X is well defined, as no two distinct inputs $\vec{x}_1 \neq \vec{x}_2$ can yield the same gradient $\vec{\nabla} G(\vec{x}_1) = \vec{\nabla} G(\vec{x}_2)$ (the restriction of G to any 1D subspace of $\mathbb{R}^\nu$, such as the one interpolating between $\vec{x}_1$ and $\vec{x}_2$, is strictly convex and so has strictly increasing derivatives; so, the gradients at these inputs can't be equal).

We can now play the same game with TG, since we know it to be strictly convex and differentiable as well. In short, let $E' = \{\vec{x} \in \mathbb{R}^\nu \ : \ \vec{\nabla}(TG)(\vec{y}) = \vec{x} \text{ for some } \vec{y} \in \mathbb{R}^\nu\}$ and define a map $Y : E' \to \mathbb{R}^\nu$ to be such that for all $\vec{x} \in E'$, we have $\vec{\nabla}(TG)(Y(\vec{x})) = \vec{x}$. With these definitions out of the way, we can prove our result.

We have that

$$(TG)(\vec{y}) = \sup_{\vec{x}} \left[\vec{y} \cdot \vec{x} - G(\vec{x})\right]$$

Since the expression inside the supremum is strictly concave as a function of $\vec{x}$, we know that this is attained precisely when the gradient of the expression is 0. This happens precisely when $\vec{\nabla} G(\vec{x}) = \vec{y} \iff \vec{x} = X(\vec{y})$. So, for all $\vec{y}$,

$$(TG)(\vec{y}) = \vec{y} \cdot X(\vec{y}) - G(X(\vec{y}))$$

Identical reasoning, replacing G with TG and $\vec{y}$ with $\vec{x}$, reveals that for all $\vec{x}$,

$$\begin{aligned}(T[TG])(\vec{x}) &= \sup_{\vec{y}} \left[\vec{x} \cdot \vec{y} - (TG)(\vec{y})\right] = \vec{x} \cdot Y(\vec{x}) - (TG)(Y(\vec{x})) \\ &= \vec{x} \cdot Y(\vec{x}) - Y(\vec{x}) \cdot X(Y(\vec{x})) + G(X(Y(\vec{x}))),\end{aligned}$$

where for the last step we simply plugged in our earlier form for TG. All that is left to do is to note that $X(Y(\vec{x})) = \vec{x}$. To see this, we can compute from the form of TG using the vector-valued function chain rule that

$$\vec{\nabla}(TG)(\vec{y}) = X(\vec{y}) + \vec{y}^T D[X(\vec{y})] - (\vec{\nabla} G)(X(\vec{y}))^T D[X(\vec{y})],$$

where $D[X(\vec{y})]$ is a $\nu \times \nu$ matrix of the componentwise derivatives of X. Note, however, that $(\vec{\nabla} G)(X(\vec{y})) = \vec{y}$ by the definition of the map X. This means that

$$\vec{\nabla}(TG)(\vec{y}) = X(\vec{y}) + \vec{y}^T D[X(\vec{y})] - \vec{y}^T D[X(\vec{y})] = X(\vec{y})$$

This means that $X(Y(\vec{x})) = \vec{x}$ by our definition of the map Y. We can plug this into our form of $T[TG]$ to see that for all $\vec{x} \in \mathbb{R}^\nu$,

$$\begin{aligned}(T[TG])(\vec{x}) &= \vec{x} \cdot Y(\vec{x}) - Y(\vec{x}) \cdot X(Y(\vec{x})) + G(X(Y(\vec{x}))) \\ &= \vec{x} \cdot Y(\vec{x}) - Y(\vec{x}) \cdot \vec{x} + G(\vec{x}) \\ &= G(\vec{x})\end{aligned}$$

Therefore, $T[TG] = G$ and we are done. ∎