## Remarks about inf-sup and min-max problems

The equality

$$\inf_{v\in D}\sup_{q\in C} S(v,q) = \sup_{q\in C}\inf_{v\in D} S(v,q).$$

is absolutely non-trivial and only holds under certain assumptions. For a counter-example consider $S(v,q) = cos(v+q)$, $D = C = \mathbb{R}$. One thing that does hold in general is

$$\begin{aligned}\inf_{v\in D}\sup_{q\in C} S(v,q) &\geq \inf_{v\in D}\sup_{q\in C}\inf_{\tilde{v}} S(\tilde{v},q)\\ &= \sup_{q\in C}\inf_{\tilde{v}\in D} S(\tilde{v},q),\\ &= \sup_{q\in C}\inf_{v\in D} S(v,q).\end{aligned}$$

For "'friendly" saddle point problems arising from "friendly" convex minimization problems the equality holds as we will later see in Fenchel's Duality Theorem.

## Fenchel-Young Inequality

By definition

$$E^*(p) = \sup_u \langle u,p\rangle - E(u),$$

such that the inequality immediately follows. Left to show is the equality statement. We have one inequality, such that we need

$$E(u) + E^*(p) \leq \langle u,p\rangle,$$

or, in other words,

$$E(u) + \langle p,z\rangle - E(z) \leq \langle u,p\rangle,\ \forall z.$$

Rewritten, the above is nothing but

$$E(z) - E(u) - \langle p, z-u\rangle \geq 0,\ \forall z,$$

or $p \in \partial E(u)$.

## Biconjugate

We'll show two proofs - one is incomplete (as it only considers the relative interior) but gives a quick intuition of why the statement makes sense. The other proof is complete but based on the separating hyperplane theorem (which we will not prove).

First of all, note that it always holds that

$$E^{**}(u) = \sup_p \langle p,u\rangle - E^*(p) \leq \sup_p \langle p,u\rangle - (\langle p,u\rangle - E(u)) = E(u),$$

by the Fenchel-Young Inequality.

**Version 1:** If $E$ is subdifferentiable at $u$, let $q \in \partial E(u)$. We readily obtain

$$E^{**}(u) = \sup_p \langle p,u\rangle - E^*(p) \geq \langle q,u\rangle - E^*(q) = E(u),$$

by the equality of the Fenchel-Young Inequality. In combination with $E^{**}(u) \leq E(u)$ as shown above, this yields $E^{**}(u) = E(u)$.

**Version 2:** We need a particular version of the separating hyperplane theorem.

**Theorem 1.** *Let $S \subset \mathbb{R}^n$ be a nonempty closed convex set, and $\mathbb{R}^n \ni u \notin S$. Then there exists a nonzero vector $z$ and a number $c < 0$ such that*

$$\langle z, v-u\rangle \leq c \qquad \forall v \in S.$$

Now let us assume that $E^{**} \neq E$. We already know that $E^{**}(u) \leq E(u)$ for all $u$, i.e. $\text{epi}(E) \subset \text{epi}(E^{**})$. Therefore, our assumption leads to $\text{epi}(E) \subsetneq \text{epi}(E^{**})$, i.e. there exists a $u$ such that $(u, E^{**}(u)) \notin \text{epi}(E)$. By the separating hyperplane theorem there exists a vector $(a, b) \in \mathbb{R}^{n+1}$ and a constant $c < 0$ such that

$$\left\langle \begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} v - u \\ \alpha - E^{**}(u) \end{pmatrix} \right\rangle \leq c < 0 \qquad \forall (v, \alpha) \in \text{epi}(E).$$

We can readily exclude $b > 0$ by choosing $v = u$ and let $\alpha$ go to infinity.

For $b = 0$ we choose $p \in \text{dom}(E^*)$ and add $\epsilon(p, -1)$ to $(a, b)$, i.e.

$$\begin{aligned}
\left\langle \begin{pmatrix} a \\ b \end{pmatrix} + \epsilon \begin{pmatrix} p \\ -1 \end{pmatrix}, \begin{pmatrix} v - u \\ \alpha - E^{**}(u) \end{pmatrix} \right\rangle &\leq c + \epsilon \left\langle \begin{pmatrix} p \\ -1 \end{pmatrix}, \begin{pmatrix} v - u \\ \alpha - E^{**}(u) \end{pmatrix} \right\rangle \\
&= c + \epsilon \left( \langle p, v \rangle - \alpha + E^{**}(u) + \langle p, u \rangle \right) \\
&\leq c + \epsilon \left( \langle p, v \rangle - E(v) + E^{**}(u) + \langle p, u \rangle \right) \\
&\leq c + \epsilon \left( E^*(p) + E^{**}(u) + \langle p, u \rangle \right) \\
&< 0 \qquad \forall (v, \alpha) \in \text{epi}(E), \text{ for } \epsilon \text{ sufficiently small.}
\end{aligned}$$

Therefore, w.r.o.g. $b < 0$.

For $b < 0$ divide the whole inequality by $-b$, define $z = a/(-b)$ and find

$$\langle z, v - u \rangle - (\alpha - E^{**}(u)) \leq c/(-b) < 0 \qquad \forall (v, \alpha) \in \text{epi}(E),$$

which means

$$\begin{aligned}
&\langle z, v \rangle - E(v) - \langle z, u \rangle + E^{**}(u) \leq c/(-b) < 0 \qquad \forall v \in \mathbb{R}^n \\
\Rightarrow &E^*(z) + E^{**}(u) - \langle z, u \rangle < 0
\end{aligned}$$

which contradicts Fenchel's inequality.

## Subgradient of convex conjugate

Let $p \in \partial E(u)$. By the Fenchel-Young Inequality we know that

$$E(u) + E^*(p) = \langle u, p \rangle.$$

On the other hand, $E = E^{**}$ such that

$$E^{**}(u) + E^*(p) = \langle u, p \rangle,$$

and the Fenchel-Young Inequality tells us that $u \in \partial E^*(p)$. Similarly, $u \in \partial E^*(p)$ implies $p \in \partial E(u)$.

## Conjugate of a strongly convex function

For a proper, closed, strongly convex function

$$\max_u \langle u, p \rangle - E(u) = -\min_u E(u) - \langle u, p \rangle$$

exists and is unique. The optimality condition immediately yields that the maximum/minimum is attained for $p \in \partial E(u)$, i.e. for $u \in \partial E^*(p)$. Since the optimal $u$ was unique, the subdifferential $\partial E^*(p)$ is single valued for all $p$, which yields the differentiability of $E^*$.

**Remark**: To see the latter in more detail, one could consider *directional derivatives*, i.e.

$$\nabla_v E(u) := \inf_{\epsilon > 0} \frac{E(u + \epsilon v) - E(u)}{\epsilon}.$$

The convexity of $E$ allows to show that the above expression is monotonically decreasing in $\epsilon$. Since we know that the expression is also bounded from below by $\langle p, v \rangle$, one can conclude the directional derivative is also equal to

$$\nabla_v E(u) = \lim_{\epsilon \to 0^+} \frac{E(u + \epsilon v) - E(u)}{\epsilon}.$$

We always get the lower bound $\nabla_v E(u) \geq \langle p, v \rangle$. If equality did not hold in the above case, one could show that the subdifferential is not single-valued.

Continuing with the proof, the convexity of $E - \frac{m}{2}\|\cdot\|^2$ yields that

$$\langle u - v, p - q \rangle \geq m\|u - v\|^2 \qquad \forall p \in \partial E(u), q \in \partial E(v),$$

or in other words

$$\langle \nabla E^*(p) - \nabla E^*(q), p - q \rangle \geq m\|\nabla E^*(p) - \nabla E^*(q)\|^2 \qquad \forall p, q,$$

which is called co-coercivity and yields the $\frac{1}{m}$-smoothness of $E^*$ by the Cauchy-Schwarz inequality.

## Fenchel's Duality Theorem

First note that the dual problem is always less or equal to the primal one (see remark at the top of this document).

Let us first do a little sanity check: Let us assume a minimum is attained at some $\hat{u}$. Then our assumptions yield that we may apply the sum rule and the optimality condition is

$$q + K^* p = 0 \qquad q \in \partial H(\hat{u}),\ p \in \partial R(K\hat{u}).$$

This implies that $u \in \partial H^*(-K^T p)$ and $Ku \in \partial R(p)$ such that

$$0 = Ku - Ku \in -K\partial H^*(-K^T p) - \partial R(p),$$

which is the optimality condition for maximizing $-H^*(-K^T p) - R(p)$.