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 do an incomplete proof (e.g. limiting ourselves to the relative interior), just to give an intuition on why the statement makes sense.

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.

If E is subdifferentiable at E, 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 $E^{**}(u) = E(u)$.

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

Fenchel's Duality Theorem

By the Fenchel-Young Inequality we know that

$$H(u) + R(Ku) \geq \langle q, u\rangle + \langle p, Ku\rangle - H^*(q) - R^*(p)$$

for all p, q. Now choose $q = -K^T p$ to obtain

$$H(u) + R(Ku) \geq -H^*(-K^T p) - R^*(p)$$

for all p. Now we know that equality holds if $q = -K^T p \in \partial H(u)$ and $p \in \partial R(Ku)$. This implies that

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

which is the optimality condition for minimizing $H(u) + R(Ku)$. On the other hand $q = -K^T p \in \partial H(u)$ and $p \in \partial R(Ku)$ means $u \in \partial H^*(-K^T p)$ 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 $-H^*(-K^T p) - R(p)$.