Strong Convexity and Smoothness Duality

Pontus Giselsson

In this short note, we prove the following duality correspondence.

Theorem 1 *The following are equivalent for* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$*.*

(i) f *is proper closed and* σ*-strongly convex*

(ii) $\partial f : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ *is maximally monotone and* σ*-strongly monotone*

(iii) $\nabla f^* : \mathbb{R}^n \to \mathbb{R}^n$ *is* σ*-cocoercive*

(iv) ∇f^* *is* $\frac{1}{\sigma}$*-Lipschitz continuous and maximally monotone*

(v) $f^* : \mathbb{R}^n \to \mathbb{R}$ *is closed convex and satisfies descent lemma (is* $\frac{1}{\sigma}$*-smooth)*

(vi) f^* *satisfies for all* $u, v \in \mathbb{R}^n$*:*

$$f^*(u) + \nabla f^*(u)^T(v-u) + \tfrac{\sigma}{2}\|\nabla f^*(v) - \nabla f^*(u)\|_2^2 \leq f^*(v). \tag{1}$$

The implication $(iv) \Rightarrow (iii)$ is called the Baillon-Haddad theorem.

We will make use of the following results.

Proposition 1 (Rockafellar) *The function* $f : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$ *is proper closed and convex if and only if* $\partial f : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ *is maximally monotone.*

Proposition 2 (Minty) *The subdifferential* $\partial f : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ *is maximally monotone if and only if* $\mathrm{ran}(\alpha I + \partial f) = \mathbb{R}^n$ *for any* $\alpha > 0$*.*

Proposition 3 *Suppose that* f *is proper closed and convex. Then* $(\partial f)^{-1} = \partial f^*$*.*

Proof. $(i) \Leftrightarrow (ii)$: (i) is equivalent to that $g(x) = f(x) - \frac{\sigma}{2}\|x\|_2^2$ is proper closed and convex and Proposition 1 implies its equivalence to that $\partial g = \partial(f - \frac{\sigma}{2}\|\cdot\|_2^2) = \partial f - \sigma I$ is maximally monotone (where the last equality can trivially be shown to hold). This, in turn, is equivalent to that ∂f is maximally monotone and σ-strongly monontone.

$(ii) \Leftrightarrow (iii)$: (ii) is equivalent to that $\partial g = \partial f - \sigma I$ is maximally monotone. The monotonicity part is equivalent to

$$(u-v)^T(x-y) \geq \sigma\|x-y\|_2^2$$

for all $(x,u) \in \mathrm{gph}\partial f$ and $(y,v) \in \mathrm{gph}\partial f$ or equivalently (Proposition 3) for all $x \in \partial f^*(u)$ and $y \in \partial f^*(v)$. Since Cauchy-Schwarz implies that ∂f^* is singlevalued on its domain $(D = \mathrm{ran}\,\partial f)$, it is equivlent to that

$$(u-v)^T(\nabla f^*(u) - \nabla f^*(v)) \geq \sigma\|\nabla f^*(u) - \nabla f^*(v)\|_2^2 \tag{2}$$

where $\nabla f^* : D \to \mathbb{R}^n$ where $D = \mathrm{ran}\partial f$.

The maximally part is (by Proposition 2) equivalent to that $\mathrm{ran}(\alpha I + \partial g) = \mathbb{R}^n$ for any $\alpha > 0$. Now set $\alpha = \sigma$ to get $\mathrm{ran}(\sigma I + \partial f - \sigma I) = \mathrm{ran}(\partial f) = D = \mathbb{R}^n$.

Hence maximal monotonicity of $g = f - \frac{\sigma}{2}\|\cdot\|_2^2$ is equivalent to that $\nabla f^* : \mathbb{R}^n \to \mathbb{R}^n$ satisfies (2), i.e., is σ-cocoercive.

$(iii) \Rightarrow (iv)$: Cauchy-Schwarz and nonnegativity of norms give that cocoercivity (2) implies monotonicity and $\frac{1}{\sigma}$-Lipschitz continuity of ∇f^*. Further, since f^* is proper closed convex (by contruction of conjugate functions) ∇f^* is maximally monotone (Proposition 1).

$(iv) \Rightarrow (v)$: Let $h(\tau) = f^*(u + \tau(v-u))$, then by chain rule

$$\nabla h(\tau) = \nabla f^*(u + \tau(v-u))^T(v-u)$$

and

$$f^*(v) - f^*(u) = h(1) - h(0) = \int_{\tau=0}^1 \nabla h(\tau)\, d\tau = \int_{\tau=0}^1 \nabla f^*(u + \tau(v-u))^T(v-u)\, d\tau.$$

Further

$$\nabla f^*(u)^T(v-u) = \int_{\tau=0}^1 \nabla f^*(u)^T(v-u)\, d\tau$$

Adding equalities on previous slide and taking absolute value:

$$\begin{aligned}
&|f^*(v) - f^*(u) - \nabla f^*(u)^T(v-u)| \\
&\quad = |\int_{\tau=0}^1 (\nabla f^*(u + \tau(v-u)) - \nabla f^*(u))^T(v-u)\, d\tau| \\
&\quad \leq \int_{\tau=0}^1 |(\nabla f^*(u + \tau(v-u)) - \nabla f^*(u))^T(v-u)|\, d\tau \\
&\quad \leq \int_{\tau=0}^1 \|\nabla f^*(u + \tau(v-u)) - \nabla f^*(u)\|_2 \|v-u\|_2\, d\tau \\
&\quad \leq \int_{\tau=0}^1 \beta\|\tau(v-u)\|_2\|v-u\|_2\, d\tau = \beta\|v-u\|_2^2 \int_{\tau=0}^1 \tau\, d\tau \\
&\quad = \tfrac{\beta}{2}\|v-u\|_2^2
\end{aligned}$$

Rearranging gives

$$\begin{aligned}
f^*(v) - f^*(u) - \nabla f^*(u)^T(v-u) &\leq \tfrac{\beta}{2}\|v-u\|_2^2 \\
f^*(v) - f^*(u) - \nabla f^*(u)^T(v-u) &\geq -\tfrac{\beta}{2}\|v-u\|_2^2.
\end{aligned}$$

Now, since f^* is closed convex, the second condition is redundant and f^* satisfies

$$\begin{aligned}
&f^*(v) - f^*(u) - \nabla f^*(u)^T(v-u) \leq \tfrac{\beta}{2}\|v-u\|_2^2 \\
&f^*(v) \geq f^*(u) + \nabla f^*(u)^T(v-u)
\end{aligned}$$

i.e., f^* is closed convex and satisfies the descent lemma.

$(v) \Rightarrow (vi)$: Define $\phi(v) = f^*(v) - \nabla f^*(u)^T v$, which is also $\frac{1}{\sigma}$-smooth (w.r.t. v) and convex with gradient: $\nabla\phi(v) = \nabla f^*(v) - \nabla f^*(u)$. A minimizing point is u since ϕ convex and $\nabla\phi(u) = 0$. Therefore, and since ϕ is smooth and the descent lemma holds, and we can conclude:

$$\begin{aligned}
\phi(u) &\leq \phi(v - \sigma\nabla\phi(v)) \leq \phi(v) + \nabla\phi(v)^T(v - \sigma\nabla\phi(v) - v) + \tfrac{1}{2\sigma}\|v - \sigma\nabla\phi(v) - v\|_2^2 \\
&= \phi(v) - \tfrac{\sigma}{2}\|\nabla\phi(v)\|_2^2.
\end{aligned}$$

Inserting the defintion of ϕ gives:

$$f^*(u) - \nabla f^*(u)^T u \leq f^*(v) - \nabla f^*(u)^T v - \tfrac{\sigma}{2}\|\nabla f^*(v) - \nabla f^*(u)\|_2^2$$

and after rearrangement

$$f^*(u) + \nabla f^*(u)^T(v-u) + \tfrac{\sigma}{2}\|\nabla f^*(v) - \nabla f^*(u)\|_2^2 \leq f^*(v),$$

which was to be proven.

$(vi) \Rightarrow (iii)$: Inequality (1) holds for arbitrary $u, v \in \mathbb{R}^n$. Adding two copies with u, v swapped gives

$$(\nabla f^*(u) - \nabla f^*(v))^T(u-v) \geq \sigma\|\nabla f^*(v) - \nabla f^*(u)\|_2^2,$$

which is the definition of cocoercivity in (iii). $\square$