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)*

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

$$(s_x - s_y)^T(x - y) \geq \sigma\|x - y\|_2^2$$

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

$$(s_x - s_y)^T(\nabla f^*(s_x) - \nabla f^*(s_y)) \geq \sigma\|\nabla f^*(s_x) - \nabla f^*(s_y)\|_2^2 \tag{1}$$

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 (1), i.e., is σ-cocoercive.

$(iii) \Rightarrow (iv)$: Cauchy-Schwarz and nonnegativity of norms give that cocoercivity (1) 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 $h(\tau) = f^*(x + \tau(y - x))$, then by chain rule $\nabla h(\tau) = \nabla f^*(x + \tau(y - x))^T(y - x)$ and

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

Further

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

Adding equalities on previous slide and taking absolute value:

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

Rearranging gives

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

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

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

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

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

Inserting the defintion of ϕ gives:

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

and after rearrangement

$$f^*(x) + \nabla f^*(x)^T(y-x) + \tfrac{\sigma}{2}\|\nabla f^*(y) - \nabla f^*(x)\|_2^2 \leq f^*(y).$$

This inequality holds for arbitrary $x, y \in \mathbb{R}^n$. We add two copies with x, y swapped to get

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

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