Lemma 2.6. Let $f : S \to \mathbb{R}$ be a convex function. Then, f is L-Lipschitz over S with respect to a norm $\|\cdot\|$ iff for all $\mathbf{w} \in S$ and $\mathbf{z} \in \partial f(\mathbf{w})$ we have that $\|\mathbf{z}\|_\star \le L$, where $\|\cdot\|_\star$ is the dual norm.

Proof. Assume that f is Lipschitz. Choose some $\mathbf{w} \in S, \mathbf{z} \in \partial f(\mathbf{w})$. Let $\mathbf{u}$ be such that $\mathbf{u} - \mathbf{w} = \operatorname{argmax}_{\mathbf{v}:\|\mathbf{v}\|=1} \langle \mathbf{v}, \mathbf{z} \rangle$. Therefore, $\langle \mathbf{u} - \mathbf{w}, \mathbf{z} \rangle = \|\mathbf{z}\|_\star$. From the definition of the sub-gradient,

$$f(\mathbf{u}) - f(\mathbf{w}) \ge \langle \mathbf{z}, \mathbf{u} - \mathbf{w} \rangle = \|\mathbf{z}\|_\star .$$

On the other hand, from the Lipschitzness of f we have

$$L = L\|\mathbf{u} - \mathbf{w}\| \ge f(\mathbf{u}) - f(\mathbf{w}) .$$

Combining the above two inequalities we conclude that $\|\mathbf{z}\|_\star \le L$. For

Assume that f is lipschitz

Choose some $x \in dom(f)$ and $g \in \partial f(x)$

Let y be such that

$$y - x = \underset{\eta : \|\eta\|_2 = 1}{\operatorname{argmax}} \ \eta^T g \qquad -①$$

Therefore,

$$(y-x)^T (g) = \|y-x\|_2 \|g\|_2$$

Since, $\|y-x\|_2 = 1$ by ①

$$(y-x)^T (g) = \|g\|_2$$

As, f is lipschitz

$$L = L\|y-x\|_2 \geq |f(y) - f(x)| \qquad -②$$

From the definition of subgradient,

$$f(y) - f(x) \geq g^T (y-x) = \|g\|_2 \qquad -③$$

Combining ② and ③, we have

$$\|g\|_2 \leq f(y) - f(x) \leq |f(y) - f(x)| \leq L\|y-x\|_2 = L$$

Therefore, $\|g\|_2 \leq L$

Cauchy-Schwartz Inequality

For any two vectors $x, y \in \mathbf{R}^n$, we have

$$x^T y \leq \|x\|_2 \cdot \|y\|_2.$$

The above inequality is an equality if and only if x, y are collinear. In other words:

$$(P) \qquad \max_{x \,:\, \|x\|_2 \leq 1} x^T y = \|y\|_2,$$

with optimal x given by $x^* = y/\|y\|_2$ if y is non-zero.

Proof: The inequality is trivial if either one of the vectors x, y is zero. Let us assume both are non-zero. Without loss of generality, we may re-scale x and assume it has unit Euclidean norm ($\|x\|_2 = 1$). Let us first prove that

$$x^T y \leq \|y\|_2.$$

We consider the polynomial

$$p(t) = \|tx - y\|_2^2 = t^2 - 2t(x^T y) + y^T y.$$

Since it is non-negative for every value of t, its discriminant $\Delta = (x^T y)^2 - y^T y$ is non-positive. The Cauchy-Schwartz inequality follows.

The second result is proven as follows. Let $v(P)$ be the optimal value of the problem. The Cauchy-Schwartz inequality implies that $v(P) \leq \|y\|_2$. To prove that the value is attained (it is equal to its upper bound), we observe that if $x = y/\|y\|_2$, then

$$x^T y = \frac{y^T y}{\|y\|_2} = \|y\|_2.$$

The vector $x = y/\|y\|_2$ is feasible for the optimization problem (P). This establishes a lower bound on the value of (P), $v(P)$:

$$\|y\|_2 \leq v(P) = \max_{x \,:\, \|x\|_2 \leq 1} x^T y.$$