Theorem 0.1. *Let f_1, f_2 be two convex functions defined on some $D \subset \mathbb{R}^n$. Then for any x we have*

$$\partial(f_1 + f_2)(x) = \partial f_1(x) + \partial f_2(x).$$

(The right-hand side is the Minkowski sum of sets $A + B = \{a + b : a \in A, b \in B\}$).

The proof is adapted from the lecture slides at the following URL (page 12):

https://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-253-convex-analysis-and-optimization-spring-2012/lecture-notes/MIT6_253S12_lec12.pdf

Proof. The inclusion $\supseteq$ is obvious. We focus on $\subseteq$. The proof will use strong duality. Let $g \in \partial(f_1 + f_2)(x)$ so that

$$f_1(y) + f_2(y) \geq f_1(x) + f_2(x) + \langle g, y - x \rangle \tag{1}$$

for all $y \in D$. Consider the following convex optimization problem

$$\min_{y_1, y_2 \in D} \quad f_1(y_1) + f_2(y_2) - \langle g, y_2 - x \rangle \quad \text{s.t.} \quad y_1 = y_2.$$

By (1), the minimum is equal to $f_1(x) + f_2(x)$. Let's formulate the dual: the Lagrangian is

$$L(y_1, y_2, \lambda) = f_1(y_1) + f_2(y_2) - \langle g, y_2 - x \rangle + \langle \lambda, y_2 - y_1 \rangle$$

and the dual function is

$$G(\lambda) = \min_{y_1, y_2} L = \left(\min_{y_1} f_1(y_1) - \langle \lambda, y_1 \rangle \right) + \left(\min_{y_2} f_2(y_2) - \langle g - \lambda, y_2 \rangle \right) + \langle g, x \rangle. \tag{2}$$

By strong duality (Slater's condition holds here, just take $y_1 = y_2$ any point in the interior of D) we know that there exists λ such that $G(\lambda) = f_1(x) + f_2(x)$. Using (2) and rearranging we get

$$\left(\min_{y_1 \in D} f_1(y_1) - (f_1(x) + \langle \lambda, y_1 - x \rangle) \right) + \left(\min_{y_2 \in D} f_2(y_2) - (f_2(x) + \langle g - \lambda, y_2 - x \rangle) \right) \ = \ 0.$$

Note that each minimum term is equal to 0: both terms are ≤ 0 by taking $y_{1,2} = x$ and since they sum to 0 both must be equal to 0.

This means that $\lambda \in \partial f_1(x)$ and $g - \lambda \in \partial f_2(x)$ as desired.

□