Old proof for the stability guarantee of `fn make_clamp`

Sílvia Casacuberta

This proof resides in **"contrib"** because it has not completed the vetting process.

Salil suggested introducing the definition of row transform and adding a general lemma for its stability guarantee, as shown in Section 3.3 in "List of definitions used in the proofs". However, we keep the longer old proof (case-by-case) for completeness, and in case any issues arise when revising the definition and theorems relating to row transforms

Proof. **(Stability guarantee).** Throughout the stability guarantee proof, we can assume that $\texttt{function}(v)$ and $\texttt{function}(w)$ are in the correct output domain, by the *appropriate output domain property* shown above.

Since by assumption $\texttt{Map}(\texttt{d\_in}) \leq \texttt{d\_out}$, by the `make_clamp` stability map, we have that $\texttt{d\_in} \leq \texttt{d\_out}$. Moreover, v, w are assumed to be `d_in`-close. By the definition of the symmetric difference metric, this is equivalent to stating that $d_{Sym}(v, w) = |\text{MultiSet}(v)\Delta\text{MultiSet}(w)| \leq \texttt{d\_in}$.

Let $\mathcal{X}$ be the domain of all elements of type `T`. By applying the histogram notation,[1] it follows that

$$d_{Sym}(v, w) = \|h_v - h_w\|_1 = \sum_{z\in\mathcal{X}} |h_v(z) - h_w(z)| \leq \texttt{d\_in} \leq \texttt{d\_out}.$$

We now consider $\text{MultiSet}(\texttt{function}(v))$ and $\text{MultiSet}(\texttt{function}(w))$. For each element $z \in \text{MultiSet}(v) \cup \text{MultiSet}(w)$, where z has type `T`, if $z \in \text{MultiSet}(v)\Delta\text{MultiSet}(w)$, we will assume wlog that $z \in \text{MultiSet}(v) \setminus \text{MultiSet}(w)$. We consider the following cases:

1. $z > \texttt{U}$ or $z < \texttt{L}$: then, in the former case, $\texttt{clamp}(z) = \texttt{U}$. First consider the case when $z \in \text{MultiSet}(v) \cup \text{MultiSet}(w)$ with the same multiplicity in both multisets. Then, $|h_{\texttt{function}(v)}(z) - h_{\texttt{function}(w)}(z)| = 0$ because we have both $h_{\texttt{function}(v)}(z) = 0$ and $h_{\texttt{function}(w)}(z) = 0$. Thus the sum

$$\sum_{z\in\mathcal{X}} |h_{\texttt{function}(v)}(z) - h_{\texttt{function}(w)}(z)|$$

remains invariant, because the quantity $|h_v(z) - h_w(z)|$ is added to $|h_{\texttt{function}(v)}(\texttt{U}) - h_{\texttt{function}(w)}(\texttt{U})|$, given that $\texttt{clamp}(z) = \texttt{U}$.

Suppose z has multiplicity $k_v \geq 0$ in $\text{MultiSet}(v)$ and multiplicity $k_w \geq 0$ in $\text{MultiSet}(w)$, where $k_v \neq k_w$. After considering z, the value $h_{\texttt{function}(v)}(\texttt{U})$ becomes $h_{\texttt{function}(v)}(\texttt{U}) + k_v$, and $h_{\texttt{function}(w)}(\texttt{U})$ becomes $h_{\texttt{function}(w)}(\texttt{U}) + k_w$. Hence the quantity $|h_{\texttt{function}(v)}(\texttt{U}) - h_{\texttt{function}(w)}(\texttt{U})|$ increases by at most $|h_v(z) - h_w(z)|$, since, by the triangle inequality,

$$\begin{aligned}
&|(h_{\texttt{function}(v)}(\texttt{U}) + k_v) - (h_{\texttt{function}(w)}(\texttt{U}) + k_w)| \leq \\
&\leq |h_{\texttt{function}(v)}(\texttt{U}) - h_{\texttt{function}(w)}(\texttt{U})| + |k_v - k_w| = \\
&= |h_{\texttt{function}(v)}(\texttt{U}) - h_{\texttt{function}(w)}(\texttt{U})| + |h_v(z) - h_w(z)|.
\end{aligned}$$

[1] Note that there is a bijection between multisets and histograms, which is why the proof can be carried out with either notion. For further details, please consult https://www.overleaf.com/project/60d214e390b337703d200982.

The same argument applies whenever $z < \texttt{L}$.[2]

2. $z \in$ `(L, U)`: then, $\texttt{clamp}(z) = z$. Since $h_v(z) = h_{\texttt{function}(v)}(z)$ and $h_v(w) = h_{\texttt{function}(w)}(z)$, it follows that $|h_v(z) - h_w(z)| = |h_{\texttt{function}(v)}(z) - h_{\texttt{function}(w)}(z)|$. Hence the histogram count, i.e., the quantity

$$\sum_{z \in \mathcal{X}} |h_{\texttt{function}(v)}(z) - h_{\texttt{function}(w)}(z)|,$$

remains invariant.

3. $z = \texttt{U}$ or $z = \texttt{L}$: in the former case, $\texttt{clamp}(z) = \texttt{U}$. If $z \in \text{MultiSet}(v) \cup \text{MultiSet}(w)$ with the same multiplicity in both multisets, then the histogram count remains invariant under the addition of element z. Otherwise, if $z \in \text{MultiSet}(v) \setminus \text{MultiSet}(w)$, or if z is in their union but with different multiplicity, then element z can increase the quantity $|h_{\texttt{function}(v)}(\texttt{U}) - h_{\texttt{function}(w)}(\texttt{U})|$ by at most $|h_v(z) - h_w(z)|$, following the same reasoning with the triangle inequality as in case 2.

   The same argument applies whenever $z = \texttt{L}$.

By aggregating the three cases above, we conclude that

$$\sum_{z \in \mathcal{X}} |h_{\texttt{function}(v)}(z) - h_{\texttt{function}(w)}(z)| \leq \sum_{z \in \mathcal{X}} |h_v(z) - h_w(z)|.$$

By the initial assumptions, we recall that $\texttt{d\_in} \leq \texttt{d\_out}$, and that v, w are `d_in`-close. Then,

$$\sum_{z \in \mathcal{X}} |h_{\texttt{function}(v)}(z) - h_{\texttt{function}(w)}(z)| \leq \sum_{z \in \mathcal{X}} |h_v(z) - h_w(z)| \leq \texttt{d\_in} \leq \texttt{d\_out}.$$

Therefore,

$$|\text{MultiSet}(\texttt{function}(v)) \Delta \text{MultiSet}(\texttt{function}(w))| \leq \texttt{d\_out},$$

as we wanted to show. □

[2]The first subcase discussed here, i.e., when $k_v = k_w$, is also proven by the triangle inequality expression above, but it seemed clean to separate the case where the total sum remains invariant.