A. Appendices

A. Proving that the objective functions are divergences

Definition A.1. Let $\mathbb{P}$ and $\mathbb{Q}$ be probability distributions and S be the set of all probability distributions with common support. A function $D : (S, S) \to \mathbb{R}_{>0}$ is a divergence if it respects the following two conditions:

$$D(\mathbb{P}, \mathbb{Q}) \geq 0$$
$$D(\mathbb{P}, \mathbb{Q}) = 0 \iff \mathbb{P} = \mathbb{Q}.$$

Definition A.2. A function f is concave on X if and only if

Lemma A.1. Let f be a concave function on X, we have that

$$\forall x_1, x_2, x_3 \in X \text{ s.t. } x_1 < x_2 \leq x_3 : \frac{f(x_3) - f(x_1)}{x_3 - x_1} \leq \frac{f(x_2) - f(x_1)}{(x_2 - x_1)}$$

and

$$\forall x_1, x_2, x_3 \in X \text{ s.t. } x_1 \leq x_2 < x_3 : \frac{f(x_3) - f(x_2)}{(x_3 - x_2)} \leq \frac{f(x_3) - f(x_1)}{x_3 - x_1}.$$

Proof. Let $\alpha = \frac{(x_3 - x_2)}{(x_3 - x_1)}$.
If $x_1 < x_2 \leq x_3$, we have that $\alpha \in [0, 1)$.
If $x_1 \leq x_2 < x_3$, we have that $\alpha \in (0, 1)$.
Either way, by concavity, we have that

$$\begin{aligned} f(x_2) &\geq \frac{(x_3 - x_2)}{(x_3 - x_1)} f(x_1) + \left(1 - \frac{(x_3 - x_2)}{(x_3 - x_1)}\right) f(x_3) \\ &= \frac{(x_3 - x_2)}{(x_3 - x_1)} f(x_1) + \frac{(x_2 - x_1)}{(x_3 - x_1)} f(x_3) \end{aligned}$$

If $x_1 < x_2 \leq x_3$, we have that:

$$\begin{aligned} f(x_2) - f(x_1) &\geq \frac{(x_1 - x_2) f(x_1) + (x_2 - x_1) f(x_3)}{(x_3 - x_1)} \\ \frac{f(x_2) - f(x_1)}{(x_2 - x_1)} &\geq \frac{f(x_3) - f(x_1)}{(x_3 - x_1)} \end{aligned}$$

If $x_1 \leq x_2 < x_3$, we have that:

$$\begin{aligned} f(x_2) - f(x_3) &\geq \frac{(x_3 - x_2) f(x_1) + (x_2 - x_3) f(x_3)}{(x_3 - x_1)} \\ \frac{f(x_2) - f(x_3)}{(x_3 - x_2)} &\geq \frac{f(x_1) - f(x_3)}{(x_3 - x_1)} \\ \frac{f(x_3) - f(x_2)}{(x_3 - x_2)} &\leq \frac{f(x_3) - f(x_1)}{(x_3 - x_1)} \end{aligned}$$

□

Lemma A.2. Let $f : \mathbb{R} \to \mathbb{R}$ be a concave function such that $f(0) = 0$. We have that

$$\forall a, b, \nabla \text{ s.t. } b \geq a > 0, \nabla \neq 0 : \frac{f(\nabla b)}{b} \leq \frac{f(\nabla a)}{a}.$$

Proof. If $\nabla > 0$ we have that $0 < \nabla a \leq \nabla b$.
By Lemma A.1 , we have that

$$\frac{f(\nabla b) - f(0)}{\nabla(b-0)} \leq \frac{f(\nabla a) - f(0)}{\nabla(a-0)}$$
$$\iff \frac{f(\nabla b)}{b} \leq \frac{f(\nabla a)}{a}$$

If $\nabla < 0$, we have that $\nabla b \leq \nabla a < 0$.
By Lemma A.1, we have that

$$\frac{f(0) - f(\nabla a)}{\nabla(0-a)} \leq \frac{f(0) - f(\nabla b)}{\nabla(0-b)}$$
$$\iff \frac{f(\nabla a)}{\nabla a} \leq \frac{f(\nabla b)}{\nabla b}$$
$$\iff \frac{f(\nabla a)}{a} \geq \frac{f(\nabla b)}{b}, \text{ since } \nabla < 0$$

Thus, when $\nabla \neq 0$, we have that

$$\frac{f(\nabla b)}{b} \leq \frac{f(\nabla a)}{a}$$

□

Lemma A.3. Let $f : \mathbb{R} \to \mathbb{R}$ be a concave function such that $f(0) = 0$, f is differentiable at 0, $f'(0) \neq 0$, $\sup_x f(x) = M > 0$, and $\arg\sup_x f(x) > 0$. Let $L(\nabla) = af(\nabla) + bf(-\nabla)$, where $a > 0$, $b > 0$, and $a \neq b$.

If $a > b$, $\exists \delta > 0$, s.t. $\forall \nabla^* \in (0, \delta) : L(\nabla^*) > 0$
If $a < b$, $\exists \delta > 0$, s.t. $\forall \nabla^* \in (-\delta, 0) : L(\nabla^*) > 0$.

Proof. By concavity, for all $\alpha \in (0, 1]$, we have $f(\alpha x^*) \geq \alpha f(x^*) > 0$.
This means that for any $\nabla \in (0, x^*]$, we have that $f(\nabla) > 0$.

By concavity, for all x, we have that $\frac{1}{2}f(x) + \frac{1}{2}f(-x) \leq f(\frac{1}{2}x - \frac{1}{2}x) = f(0) = 0$.
Thus, for all $\nabla \in (0, x^*]$ we have that $0 < f(\nabla) \leq -f(-\nabla)$.
This means that $f(\nabla) > 0$ and $f(-\nabla) < 0$.

Let $R(x) = \frac{g(x)}{f(x)}$, where $g(x) = -f(-x)$.
We can show that:

$$\lim_{x \to 0} R(x) = \lim_{x \to 0} \frac{g(x)}{f(x)} \overset{\text{H}}{=} \lim_{x \to 0} \frac{g'(x)}{f'(x)} = \lim_{x \to 0} \frac{f'(-x)}{f'(x)} = \frac{f'(0)}{f'(0)} = 1.$$

If $\nabla \in (0, x^*]$, by concavity we have that $0 < f(\nabla) \leq -f(-\nabla)$, thus $R(\nabla) = \frac{-f(-\nabla)}{f(\nabla)} \geq 1$.

Let $\epsilon = \frac{(a'-b')}{b'}$, where $a' > b' > 0$.
By the definition of the limit, $\exists \delta > 0$ s.t. $\forall x$ s.t. $0 < |x| < \delta$, we have

$$|R(x) - 1| < \epsilon.$$

Since this is true for all x s.t. $0 < |x| < \delta$, this is also true for all $0 < \nabla^* < \min(x^*, \delta)$.

This means that

$$
\begin{aligned}
& |R(\nabla^*) - 1| < \epsilon \\
\implies & (R(\nabla^*) - 1) < \frac{(a' - b')}{b'}, \text{ since } R(\nabla) \geq 1 \text{ for all } \nabla \in (0, x^*] \\
\implies & R(\nabla^*) < \frac{a'}{b'} \\
\implies & \frac{-f(-\nabla^*)}{f(\nabla^*)} < \frac{a'}{b'} \\
\implies & a' f(\nabla^*) + b' f(-\nabla^*) > 0
\end{aligned}
$$

If $a > b$, let $a' = a$, $b' = b$, and we have $af(\nabla^*) + bf(-\nabla^*) > 0$ for all $0 < \nabla^* < \min(x^*, \delta)$.
If $a < b$, let $a' = b$, $b' = a$, and we have $af(\nabla^*) + bf(-\nabla^*) > 0$ for all $-\min(x^*, \delta) < \nabla^* < 0$.

□

Theorem A.4. Let $f : \mathbb{R} \to \mathbb{R}$ be a concave function such that $f(0) = 0$, f is differentiable at 0, $f'(0) \neq 0$, $\sup_x f(x) = M > 0$, and $\arg\sup_x f(x) > 0$. Let $\mathbb{P}$ and $\mathbb{Q}$ be probability distributions with support $\mathcal{X}$. Then, we have that

$$
\mathrm{D}_f^{Rp}(\mathbb{P}, \mathbb{Q}) = \sup_{C:\mathcal{X}\to\mathbb{R}} \mathop{\mathbb{E}}_{\substack{x\sim\mathbb{P} \\ y\sim\mathbb{Q}}} \left[f\left(C(x) - C(y) \right) \right]
$$

is a divergence.

Proof. Let $C^w(x) = k \ \forall x$ (worst possible choice of C).
Let $C^*(x) = \arg\sup_{C:\mathcal{X}\to\mathbb{R}} \mathop{\mathbb{E}}_{\substack{x\sim\mathbb{P} \\ y\sim\mathbb{Q}}} \left[f\left(C(x) - C(y) \right) \right]$ (best possible choice of C).

#1 Proof that $\mathrm{D}_f^{Rp}(\mathbb{P}, \mathbb{Q}) \geq 0$

$$
\mathrm{D}_f^{Rp}(\mathbb{P}, \mathbb{Q}) = \mathop{\mathbb{E}}_{\substack{x\sim\mathbb{P} \\ y\sim\mathbb{Q}}} \left[f\left(C^*(x) - C^*(y) \right) \right] \geq \mathop{\mathbb{E}}_{\substack{x\sim\mathbb{P} \\ y\sim\mathbb{Q}}} \left[f\left(C^w(x) - C^w(y) \right) \right] = 0.
$$

#2 Proof that $\mathbb{P} = \mathbb{Q} \implies \mathrm{D}_f^{Rp}(\mathbb{P}, \mathbb{Q}) = 0$

$$
\begin{aligned}
\mathrm{D}_f^{Rp}(\mathbb{P}, \mathbb{Q}) &= \mathop{\mathbb{E}}_{\substack{x\sim\mathbb{P} \\ y\sim\mathbb{P}}} \left[f\left(C^*(x) - C^*(y) \right) \right] \\
&= \mathop{\mathbb{E}}_{x\sim\mathbb{P}} \left[\mathop{\mathbb{E}}_{y\sim\mathbb{P}} \left[f\left(C^*(x) - C^*(y) \right) | x \right] \right] \\
&\leq \mathop{\mathbb{E}}_{x\sim\mathbb{P}} \left[f\left(\mathop{\mathbb{E}}_{y\sim\mathbb{P}} \left[C^*(x) - C^*(y) | x \right] \right) \right] \\
&= \mathop{\mathbb{E}}_{x\sim\mathbb{P}} \left[f\left(C^*(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{P}} \left[C^*(y) \right] \right) \right] \\
&= \mathop{\mathbb{E}}_{x\sim\mathbb{P}} \left[f\left(C'^*(x) \right) \right], \text{ where } C'^*(x) = C^*(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{P}} \left[C^*(y) \right] \\
&\leq f\left(\mathop{\mathbb{E}}_{x\sim\mathbb{P}} \left[C'^*(x) \right] \right), \text{ by Jensen's inequality} \\
&= f(0) \\
&= 0
\end{aligned}
$$

Since $\mathrm{D}_f^{Rp}(\mathbb{P},\mathbb{Q}) \geq 0$, we have that $\mathrm{D}_f^{Rp}(\mathbb{P},\mathbb{Q}) = 0$.

#3 Proof that $\mathrm{D}_f^{Rp}(\mathbb{P},\mathbb{Q}) = 0 \implies \mathbb{P} = \mathbb{Q}$

We prove this by contraposition (i.e., we prove that $\mathbb{P} \neq \mathbb{Q} \implies \mathrm{D}_f^{Rp}(\mathbb{P},\mathbb{Q}) \neq 0$). To do so, we design a function C' that is better than the worse option ($C(x) = k \ \forall x$).

Assume that $\mathbb{P} \neq \mathbb{Q}$.

Let $T = \arg\sup_S \mathbb{P}(S) - \mathbb{Q}(S)$ [3].
Let $p = \int_T d\mathbb{P}(x) \implies (1-p) = \int_{\mathcal{X}\setminus T} d\mathbb{P}(x)$.
Let $q = \int_T d\mathbb{Q}(y) \implies (1-q) = \int_{\mathcal{X}\setminus T} d\mathbb{Q}(y)$.
Since $\mathbb{P} \neq \mathbb{Q}$, we know that $T \neq \varnothing$.
This means that $p > 0$, $q > 0$, and $p > q$.

Let $C'(x) = \begin{cases} \nabla & \text{if } x \in T \\ 0 & \text{else} \end{cases}$, where $\nabla \neq 0$.
Let $L(\nabla) = \underset{\substack{x\sim\mathbb{P} \\ y\sim\mathbb{Q}}}{\mathbb{E}} \left[f\left(C'(x) - C'(y)\right) \right]$.

We have that

$$
\begin{aligned}
L(\nabla) &= \int_{\mathcal{X}} \int_{\mathcal{X}} f\left(C'(x) - C'(y)\right) d\mathbb{P}(x) d\mathbb{Q}(y) \\
&= \int_T \int_T f\left(C'(x) - C'(y)\right) d\mathbb{P}(x) d\mathbb{Q}(y) + \int_T \int_{\mathcal{X}\setminus T} f\left(C'(x) - C'(y)\right) d\mathbb{P}(x) d\mathbb{Q}(y) + \\
&\quad \int_{\mathcal{X}\setminus T} \int_T f\left(C'(x) - C'(y)\right) d\mathbb{P}(x) d\mathbb{Q}(y) + \int_{\mathcal{X}\setminus T} \int_{\mathcal{X}\setminus T} f\left(C'(x) - C'(y)\right) d\mathbb{P}(x) d\mathbb{Q}(y) \\
&\quad = (1) + (2) + (3) + (4)
\end{aligned}
$$

$$
\begin{aligned}
&(1) \int_T \int_T f\left(C'(x) - C'(y)\right) d\mathbb{P}(x) d\mathbb{Q}(y) = \int_T \int_T f\left(\nabla - \nabla\right) d\mathbb{P}(x) d\mathbb{Q}(y) = 0 \\
&(2) \int_T \int_{\mathcal{X}\setminus T} f\left(C'(x) - C'(y)\right) d\mathbb{P}(x) d\mathbb{Q}(y) = f(\nabla) \int_T d\mathbb{P}(x) \int_{\mathcal{X}\setminus T} d\mathbb{Q}(y) = f(\nabla) p(1-q) \\
&(3) \int_{\mathcal{X}\setminus T} \int_T f\left(C'(x) - C'(y)\right) d\mathbb{P}(x) d\mathbb{Q}(y) = f(-\nabla) \int_{\mathcal{X}\setminus T} d\mathbb{P}(x) \int_T d\mathbb{Q}(y) = f(-\nabla) q(1-p) \\
&(4) \int_{\mathcal{X}\setminus T} \int_{\mathcal{X}\setminus T} f\left(C'(x) - C'(y)\right) d\mathbb{P}(x) d\mathbb{Q}(y) = \int_{\mathcal{X}\setminus T} \int_{\mathcal{X}\setminus T} f\left(0 - 0\right) d\mathbb{P}(x) d\mathbb{Q}(y) = 0
\end{aligned}
$$

This means that $L(\nabla) = af(\nabla) + bf(-\nabla)$, where $a = p(1-q) > 0$ and $b = q(1-p) > 0$.
We know that $a = p(1-q) > q(1-p) = b$.
Thus, by Lemma A.4, we have that $\exists \nabla^* > 0$ s.t. $L(\nabla^*) > 0$.

Thus, if we let $\nabla = \nabla^*$, we have that
$\mathrm{D}_f^{Rp}(\mathbb{P},\mathbb{Q}) = \underset{\substack{x\sim\mathbb{P} \\ y\sim\mathbb{Q}}}{\mathbb{E}} \left[f\left(C^*(x) - C^*(y)\right) \right] \geq \underset{\substack{x\sim\mathbb{P} \\ y\sim\mathbb{Q}}}{\mathbb{E}} \left[f\left(C'(x) - C'(y)\right) \right] > 0.$

$\square$

Theorem A.5. Let $f : \mathbb{R} \to \mathbb{R}$ be a concave function such that $f(0) = 0$, f is differentiable at 0, $f'(0) \neq 0$, $\sup_x f(x) =$

[3] If $\mathbb{P}$ and $\mathbb{Q}$ have probability density functions $p(x)$ and $q(x)$ respectively, then $T = \{x | p(x) > q(x)\}$.

$M > 0$, and $\arg\sup_x f(x) > 0$. Let $\mathbb{P}$ and $\mathbb{Q}$ be probability distributions with support $\mathcal{X}$. Then, we have that

$$\mathrm{D}_f^{Ralf}(\mathbb{P},\mathbb{Q}) = \sup_{C:\mathcal{X}\to\mathbb{R}} \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{Q}} C(y)\right)\right]$$

is a divergence.

Proof. Let $C^w(x) = k\ \forall x$ (worst possible choice of C).
Let $C^*(x) = \arg\sup_{C:\mathcal{X}\to\mathbb{R}} \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{Q}} C(y)\right)\right]$ (best possible choice of C).

#1 Proof that $\mathrm{D}_f^{Ralf}(\mathbb{P},\mathbb{Q}) \geq 0$

$$\begin{aligned}
\mathrm{D}_f^{Ralf}(\mathbb{P},\mathbb{Q}) &= \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C^*(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{Q}} C^*(y)\right)\right] \\
&\geq \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C^w(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{Q}} C^w(y)\right)\right] \\
&= \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(k-k\right)\right] \\
&= 0.
\end{aligned}$$

#2 Proof that $\mathbb{P} = \mathbb{Q} \implies \mathrm{D}_f^{Ralf}(\mathbb{P},\mathbb{Q}) = 0$

$$\begin{aligned}
\mathrm{D}_f^{Ralf}(\mathbb{P},\mathbb{Q}) &= \sup_{C:\mathcal{X}\to\mathbb{R}} \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{Q}} C(y)\right)\right] \\
&= \sup_{C:\mathcal{X}\to\mathbb{R}} \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{P}} C(y)\right)\right], \text{ since } \mathbb{P}=\mathbb{Q} \\
&= \sup_{\substack{C':\mathcal{X}\to\mathbb{R} \\ \text{s.t. } \mathbb{E}[C'(x)]=0}} \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C'(x)\right)\right] \\
&= \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C'^*(x)\right)\right], \text{ where } C'^* = \arg\sup_{\substack{C':\mathcal{X}\to\mathbb{R} \\ \text{s.t. } \mathbb{E}[C'(x)]=0}} \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C'(x)\right)\right] \\
&\leq f\left(\mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[C'^*(x)\right]\right), \text{ by Jensen's inequality} \\
&= f(0) \\
&= 0
\end{aligned}$$

Since $\mathrm{D}_f^{Ralf}(\mathbb{P},\mathbb{Q}) \geq 0$, we have that $\mathrm{D}_f^{Ralf}(\mathbb{P},\mathbb{Q}) = 0$.

#3 Proof that $\mathrm{D}_f^{Ra}(\mathbb{P},\mathbb{Q}) = 0 \implies \mathbb{P} = \mathbb{Q}$

We prove this by contraposition (i.e., we prove that $\mathbb{P} \neq \mathbb{Q} \implies \mathrm{D}_f^{Ra}(\mathbb{P},\mathbb{Q}) \neq 0$). To do so, we design a function C' that is better than the worse option ($C(x) = k\ \forall x$).

Assume that $\mathbb{P} \neq \mathbb{Q}$.

Let $T = \arg\sup_S \mathbb{P}(S) - \mathbb{Q}(S)$.
Let $p = \int_T d\mathbb{P}(x) \implies (1-p) = \int_{\mathcal{X}\setminus T} d\mathbb{P}(x)$.
Let $q = \int_T d\mathbb{Q}(y) \implies (1-q) = \int_{\mathcal{X}\setminus T} d\mathbb{Q}(y)$.

Since $\mathbb{P} \neq \mathbb{Q}$, we know that $T \neq \varnothing$.
This means that $p > 0$, $q > 0$, and $p > q$.

Let $C'(x) = \begin{cases} \nabla & \text{if } x \in T \\ 0 & \text{else} \end{cases}$, where $\nabla \neq 0$.

Let $L(\nabla) = \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C'(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{Q}} C'(y) \right) \right]$.

We have that

$$\begin{aligned} L(\nabla) &= \int_{\mathcal{X}} f\left(C'(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{Q}} C'(y) \right) d\mathbb{P}(x) \\ &= \int_{\mathcal{X}} f\left(C'(x) - \int_T \nabla d\mathbb{Q}(y) \right) d\mathbb{P}(x) \\ &= \int_{\mathcal{X}} f\left(C'(x) - \nabla q \right) d\mathbb{P}(x) \\ &= \int_T f\left(\nabla - \nabla q \right) d\mathbb{P}(x) + \int_{\mathcal{X}\setminus T} f\left(0 - \nabla q \right) d\mathbb{P}(x) \\ &= pf\left(\nabla(1-q) \right) + (1-p) f\left(-\nabla q \right) \end{aligned}$$

Case 1: If $q < (1-q)$, by Lemma A.3, we have that:

$$\begin{aligned} \frac{f(-\nabla(1-q))}{(1-q)} &\leq \frac{f(-\nabla q)}{q} \\ \implies f(-\nabla q) &\geq \frac{q}{(1-q)} f(-\nabla(1-q)) \end{aligned}$$

Thus, $L(\nabla) \geq pf\left(\nabla(1-q)\right) + \frac{(1-p)q}{(1-q)} f\left(-\nabla(1-q)\right)$.
Knowing that $p > q$ and $(1-p) < (1-q)$, we have that $p > q > \frac{q(1-p)}{(1-q)}$.
Thus, by Lemma A.4, we have that $\exists \nabla^* > 0$ s.t. $L(\nabla^*) > 0$.

Case 2: If $q \geq (1-q)$, by Lemma A.3, we have that:

$$\begin{aligned} \frac{f(\nabla q)}{q} &\leq \frac{f(\nabla(1-q))}{(1-q)} \\ \implies f(\nabla(1-q)) &\geq \frac{(1-q)}{q} f(\nabla q) \end{aligned}$$

Thus, $L(\nabla) \geq \frac{p(1-q)}{q} f\left(\nabla q\right) + (1-p) f\left(-\nabla q\right)$.
Knowing that $p > q$ and $(1-p) < (1-q)$, we have that $(1-p) < (1-q) < \frac{(1-q)p}{q}$.
Thus, by Lemma A.4, we have that $\exists \nabla^* > 0$ s.t. $L(\nabla^*) > 0$.

Thus, if we let $\nabla = \nabla^*$, we have that
$\mathrm{D}_f^{Ralf}(\mathbb{P},\mathbb{Q}) = \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C^*(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{Q}} C^*(y) \right) \right] \geq \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C'(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{Q}} C'(y) \right) \right] > 0.$

□

Theorem A.6. Let $f : \mathbb{R} \to \mathbb{R}$ be a concave function such that $f(0) = 0$, f is differentiable at 0, $f'(0) \neq 0$, $\sup_x f(x) = M > 0$, and $\arg\sup_x f(x) > 0$. Let $\mathbb{P}$ and $\mathbb{Q}$ be probability distributions with support $\mathcal{X}$. Then, we have that

$$\mathrm{D}_f^{Ra}(\mathbb{P},\mathbb{Q}) = \sup_{C:\mathcal{X}\to\mathbb{R}} \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{Q}} C(y) \right) \right] + \mathop{\mathbb{E}}_{y\sim\mathbb{Q}}\left[f\left(\mathop{\mathbb{E}}_{x\sim\mathbb{P}} C(x) - C(y) \right) \right]$$

is a divergence.

Proof. Let $C^w(x) = k \ \forall x$ (worst possible choice of C).
Let $C^*(x) = \underset{C:\mathcal{X}\to\mathbb{R}}{\arg\sup} \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(C(x) - \underset{y\sim\mathbb{Q}}{\mathbb{E}} C(y)\right)\right] + \underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[f\left(\underset{x\sim\mathbb{P}}{\mathbb{E}} C(x) - C(y)\right)\right]$
(best possible choice of C).

#1 Proof that $D_f^{Ra}(\mathbb{P},\mathbb{Q}) \geq 0$

$$\begin{aligned}
D_f^{Ra}(\mathbb{P},\mathbb{Q}) &= \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(C^*(x) - \underset{y\sim\mathbb{Q}}{\mathbb{E}} C^*(y)\right)\right] + \underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[f\left(\underset{x\sim\mathbb{P}}{\mathbb{E}} C^*(x) - C^*(y)\right)\right] \\
&\geq \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(C^w(x) - \underset{y\sim\mathbb{Q}}{\mathbb{E}} C^w(y)\right)\right] + \underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[f\left(\underset{x\sim\mathbb{P}}{\mathbb{E}} C^w(x) - C^w(y)\right)\right] \\
&= \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(k-k\right)\right] + \underset{x\sim\mathbb{Q}}{\mathbb{E}}\left[f\left(k-k\right)\right] \\
&= 0.
\end{aligned}$$

#2 Proof that $\mathbb{P} = \mathbb{Q} \implies D_f^{Ra}(\mathbb{P},\mathbb{Q}) = 0$

Let $C'(x) = C(x) - \underset{x\sim\mathbb{P}}{\mathbb{E}} C(x)$

$$\begin{aligned}
D_f^{Ra}(\mathbb{P},\mathbb{Q}) &= \sup_{C:\mathcal{X}\to\mathbb{R}} \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(C(x) - \underset{y\sim\mathbb{Q}}{\mathbb{E}} C(y)\right)\right] + \underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[f\left(\underset{x\sim\mathbb{P}}{\mathbb{E}} C(x) - C(y)\right)\right] \\
&= \sup_{C:\mathcal{X}\to\mathbb{R}} \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(C(x) - \underset{y\sim\mathbb{P}}{\mathbb{E}} C(y)\right)\right] + \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(\underset{x\sim\mathbb{P}}{\mathbb{E}} C(y) - C(x)\right)\right] \\
&= \sup_{\substack{C':\mathcal{X}\to\mathbb{R} \\ \text{s.t. } \mathbb{E}[C'(x)]=0}} \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(C'(x)\right) + f\left(-C'(x)\right)\right] \\
&\leq 2 \sup_{\substack{C':\mathcal{X}\to\mathbb{R} \\ \text{s.t. } \mathbb{E}[C'(x)]=0}} \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(\frac{1}{2}C'(x) - \frac{1}{2}C'(x)\right)\right], \text{ by concavity} \\
&= 2 \sup_{\substack{C':\mathcal{X}\to\mathbb{R} \\ \text{s.t. } \mathbb{E}[C'(x)]=0}} \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(0\right)\right] \\
&= 0
\end{aligned}$$

Since $D_f^{Ra}(\mathbb{P},\mathbb{Q}) \geq 0$, we have that $D_f^{Ra}(\mathbb{P},\mathbb{Q}) = 0$.

#3 Proof that $D_f^{Ra}(\mathbb{P},\mathbb{Q}) = 0 \implies \mathbb{P} = \mathbb{Q}$

We prove this by contraposition (i.e., we prove that $\mathbb{P} \neq \mathbb{Q} \implies D_f^{Ra}(\mathbb{P},\mathbb{Q}) \neq 0$). To do so, we design a function C' that is better than the worse option ($C(x) = k \ \forall x$).

Assume that $\mathbb{P} \neq \mathbb{Q}$.

Let $T = \arg\sup_S \mathbb{P}(S) - \mathbb{Q}(S)$.
Let $p = \int_T d\mathbb{P}(x) \implies (1-p) = \int_{\mathcal{X}\setminus T} d\mathbb{P}(x)$.
Let $q = \int_T d\mathbb{Q}(y) \implies (1-q) = \int_{\mathcal{X}\setminus T} d\mathbb{Q}(y)$.

Since $\mathbb{P} \neq \mathbb{Q}$, we know that $T \neq \varnothing$.
This means that $p > 0$, $q > 0$, and $p > q$.

Let $C'(x) = \begin{cases} \nabla & \text{if } x \in T \\ 0 & \text{else} \end{cases}$, where $\nabla \neq 0$.
Let $L(\nabla) = \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(C'(x) - \underset{y\sim\mathbb{Q}}{\mathbb{E}} C'(y)\right)\right] + \underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[f\left(\underset{x\sim\mathbb{P}}{\mathbb{E}} C'(x) - C'(y)\right)\right]$.

We have that

$$
\begin{aligned}
L(\nabla) &= \int_{\mathcal{X}} f\left(C'(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{Q}} C'(y)\right) d\mathbb{P}(x) + \int_{\mathcal{X}} f\left(\mathop{\mathbb{E}}_{x\sim\mathbb{P}} C'(x) - C'(y)\right) d\mathbb{Q}(y) \\
&= \int_{\mathcal{X}} f\left(C'(x) - \int_T \nabla d\mathbb{Q}(y)\right) d\mathbb{P}(x) + \int_{\mathcal{X}} f\left(\int_T \nabla d\mathbb{P}(x) - C'(y)\right) d\mathbb{Q}(y) \\
&= \int_{\mathcal{X}} f\left(C'(x) - \nabla q\right) d\mathbb{P}(x) + \int_{\mathcal{X}} f\left(\nabla p - C'(y)\right) d\mathbb{Q}(y) \\
&= \int_T f\left(\nabla(1-q)\right) d\mathbb{P}(x) + \int_{\mathcal{X}\setminus T} f\left(-\nabla q\right) d\mathbb{P}(x) + \\
&\quad \int_T f\left(\nabla(p-1)\right) d\mathbb{Q}(y) + \int_T f\left(\nabla p\right) d\mathbb{Q}(y) \\
&= pf\left(\nabla(1-q)\right) + (1-p)f\left(-\nabla q\right) + qf\left(\nabla(p-1)\right) + (1-q)f\left(\nabla p\right) \\
&= pf\left(\nabla(1-q)\right) + (1-p)f\left(-\nabla q\right) + qf\left(-\nabla(1-p)\right) + (1-q)f\left(\nabla p\right)
\end{aligned}
$$

Case 1: If $(1-q) \geq p$, by Lemma A.3, we have that:

$$
\begin{aligned}
&\frac{f(\nabla(1-q))}{(1-q)} \leq \frac{f(\nabla p)}{p} \\
\implies &f(\nabla p) \geq \frac{p}{(1-q)} f(\nabla(1-q))
\end{aligned}
$$

Also, we have that $(1-p) \geq q$, thus, by Lemma A.3, we have that:

$$
\begin{aligned}
&\frac{f(-\nabla(1-p))}{(1-p)} \leq \frac{f(-\nabla q)}{q} \\
\implies &f(-\nabla q) \geq \frac{q}{(1-p)} f(-\nabla(1-p))
\end{aligned}
$$

Also, $q < p \implies (1-q) > (1-p)$, thus, by Lemma A.3, we have that:

$$
\begin{aligned}
&\frac{f(-\nabla(1-q))}{(1-q)} \leq \frac{f(-\nabla(1-p))}{(1-p)} \\
\implies &f(-\nabla(1-p)) \geq \frac{(1-p)}{(1-q)} f(-\nabla(1-q))
\end{aligned}
$$

Thus,

$$
\begin{aligned}
L(\nabla) &= pf\left(\nabla(1-q)\right) + (1-p)f\left(-\nabla q\right) + qf\left(-\nabla(1-p)\right) + (1-q)f\left(\nabla p\right) \\
&\geq pf\left(\nabla(1-q)\right) + qf\left(-\nabla(1-p)\right) + qf\left(-\nabla(1-p)\right) + pf\left(\nabla(1-q)\right) \\
&= 2pf\left(\nabla(1-q)\right) + 2qf\left(-\nabla(1-p)\right) \\
&\geq 2pf\left(\nabla(1-q)\right) + 2\frac{q(1-p)}{(1-q)} f\left(-\nabla(1-q)\right)
\end{aligned}
$$

Knowing that $p > q$ and $(1-p) < (1-q)$, we have that $2p > 2q > \frac{2q(1-p)}{(1-q)}$.
Thus, by Lemma A.4, we have that $\exists \nabla^* > 0$ s.t. $L(\nabla^*) > 0$.

Case 2: If $p > (1-q)$, by Lemma A.3, we have that:

$$
\begin{aligned}
&\frac{f(\nabla p)}{p} \leq \frac{f(\nabla(1-q))}{(1-q)} \\
\implies &f(\nabla(1-q)) \geq \frac{(1-q)}{p} f(\nabla p)
\end{aligned}
$$

Also, we have that $q > (1-p)$, thus, by Lemma A.3, we have that:

$$\frac{f(-\nabla q)}{q} \leq \frac{f(-\nabla(1-p))}{(1-p)}$$
$$\implies f(-\nabla(1-p)) \geq \frac{(1-p)}{q} f(-\nabla q)$$

Also, $p > q$, thus, by Lemma A.3, we have that:

$$\frac{f(-\nabla p)}{p} \leq \frac{f(-\nabla q)}{q}$$
$$\implies f(-\nabla q) \geq \frac{q}{p} f(-\nabla p)$$

Thus,

$$\begin{aligned} L(\nabla) = &\ pf\left(\nabla(1-q)\right) + (1-p)f\left(-\nabla q\right) + qf\left(-\nabla(1-p)\right) + (1-q)f\left(\nabla p\right) \\ \geq &\ (1-q)f\left(\nabla p\right) + (1-p)f\left(-\nabla q\right) + (1-p)f\left(-\nabla q\right) + (1-q)f\left(\nabla p\right) \\ = &\ 2(1-q)f\left(\nabla p\right) + 2(1-p)f\left(-\nabla q\right) \\ \geq &\ 2(1-q)f\left(\nabla p\right) + 2\frac{q(1-p)}{p} f\left(-\nabla p\right) \end{aligned}$$

Knowing that $p > q$ and $(1-p) < (1-q)$, we have that $2(1-q) > 2(1-p) > 2\frac{q(1-p)}{p}$.
Thus, by Lemma A.4, we have that $\exists \nabla^* > 0$ s.t. $L(\nabla^*) > 0$.

Thus, if we let $\nabla = \nabla^*$, we have that

$$\begin{aligned} \mathrm{D}_f^{Ra}(\mathbb{P},\mathbb{Q}) = &\ \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C^*(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{Q}} C^*(y)\right)\right] + \mathop{\mathbb{E}}_{y\sim\mathbb{Q}}\left[f\left(\mathop{\mathbb{E}}_{x\sim\mathbb{P}} C^*(x) - C^*(y)\right)\right] \\ \geq &\ \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C'(x) - \mathop{\mathbb{E}}_{y\sim\mathbb{Q}} C'(y)\right)\right] + \mathop{\mathbb{E}}_{y\sim\mathbb{Q}}\left[f\left(\mathop{\mathbb{E}}_{x\sim\mathbb{P}} C'(x) - C'(y)\right)\right] \\ > &\ 0. \end{aligned}$$

□

Theorem A.7. Let $f : \mathbb{R} \to \mathbb{R}$ be a concave function such that $f(0) = 0$, f is differentiable at 0, $f'(0) \neq 0$, $\sup_x f(x) = M > 0$, and $\arg\sup_x f(x) > 0$. Let $\mathbb{P}$ and $\mathbb{Q}$ be probability distributions with support $\mathcal{X}$. Let $\mathbb{M} = \frac{1}{2}\mathbb{P} + \frac{1}{2}\mathbb{Q}$ Then, we have that

$$\mathrm{D}_f^{Rc}(\mathbb{P},\mathbb{Q}) = \sup_{C:\mathcal{X}\to\mathbb{R}} \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C(x) - \mathop{\mathbb{E}}_{m\sim\mathbb{M}} C(m)\right)\right] + \mathop{\mathbb{E}}_{y\sim\mathbb{Q}}\left[f\left(\mathop{\mathbb{E}}_{m\sim\mathbb{M}} C(m) - C(y)\right)\right]$$

is a divergence.

Proof. Let $C^w(x) = k\ \forall x$ (worst possible choice of C).
Let $C^*(x) = \arg\sup_{C:\mathcal{X}\to\mathbb{R}} \mathop{\mathbb{E}}_{x\sim\mathbb{P}}\left[f\left(C(x) - \mathop{\mathbb{E}}_{m\sim\mathbb{M}} C(m)\right)\right] + \mathop{\mathbb{E}}_{y\sim\mathbb{Q}}\left[f\left(\mathop{\mathbb{E}}_{m\sim\mathbb{M}} C(m) - C(y)\right)\right]$
(best possible choice of C).

#1 Proof that $\mathrm{D}_f^{Rc}(\mathbb{P},\mathbb{Q}) \geq 0$

Same proof as theorem A.6 #1.

#2 Proof that $\mathbb{P} = \mathbb{Q} \implies \mathrm{D}_f^{Rc}(\mathbb{P},\mathbb{Q}) = 0$

Same proof as theorem A.6 #2.

#3 Proof that $\mathrm{D}_f^{Rc}(\mathbb{P},\mathbb{Q}) = 0 \implies \mathbb{P} = \mathbb{Q}$

We prove this by contraposition (i.e., we prove that $\mathbb{P} \neq \mathbb{Q} \implies \mathrm{D}_f^{Rc}(\mathbb{P}, \mathbb{Q}) \neq 0$). To do so, we design a function C' that is better than the worse option ($C(x) = k \ \forall x$).

Assume that $\mathbb{P} \neq \mathbb{Q}$.

Make the same assumptions as theorem A.6 #2. The only thing that changes is $L(\nabla)$.

We instead have that

$$\begin{aligned} L(\nabla) = & \ pf\left(\nabla(1-c)\right) + (1-p)f\left(-\nabla c\right) + qf\left(-\nabla(1-c)\right) + (1-q)f\left(\nabla c\right) \\ = & \ L_1(\nabla) + L_2(\nabla), \end{aligned}$$

where $c = \frac{1}{2}p + \frac{1}{2}q$,
$L_1(\nabla) = pf\left(\nabla(1-c)\right) + qf\left(-\nabla(1-c)\right)$,
$L_2(\nabla) = (1-q)f\left(\nabla c\right) + (1-p)f\left(-\nabla c\right)$.

Knowing that $p > q$ and $(1-q) > (1-p)$, we can use Lemma A.4 to show that
$\exists \delta_1 > 0$, s.t. $\forall \nabla_1^* \in (0, \delta_1) : L_1(\nabla_1^*) > 0$ and $\exists \delta_2 > 0$, s.t. $\forall \nabla_2^* \in (0, \delta_2) : L_2(\nabla_2^*) > 0$.
Thus, let $\delta = \min(\delta_1, \delta_2)$. We have that $\forall \nabla^* \in (0, \delta) : L_1(\nabla^*) > 0$ and $L_2(\nabla^*) > 0$.
This means that $L(\nabla) = L_1(\nabla^*) + L_2(\nabla^*) > 0$

□

B. Inequalities between Relativistic Divergences

To prove that D_1 is weaker than D_2, we can just show that $\mathrm{D}_1(\mathbb{P}, \mathbb{Q}) \leq \mathrm{D}_2(\mathbb{P}, \mathbb{Q})$ since we have that:

$$\mathrm{D}_1(\mathbb{P}_n, \mathbb{P}) \leq \mathrm{D}_2(\mathbb{P}_n, \mathbb{P}) \to 0 \implies \mathrm{D}_1(\mathbb{P}_n, \mathbb{P}) \to 0.$$

Theorem A.8. Let $f : \mathbb{R} \to \mathbb{R}$ be a concave function such that $f(0) = 0$, f is differentiable at 0, $f'(0) \neq 0$, $\sup_x f(x) = M > 0$, and $\arg\sup_x f(x) > 0$. Let $\mathbb{P}$ and $\mathbb{Q}$ be probability distributions with support $\mathcal{X}$. Then, we have that

- $\mathrm{D}^S(\mathbb{P}, \mathbb{Q}) \leq \mathrm{D}_f^{Rp}(\mathbb{P}, \mathbb{Q})$

- $\mathrm{D}_f^{Rp}(\mathbb{P}, \mathbb{Q}) \leq \mathrm{D}_f^{Ralf}(\mathbb{P}, \mathbb{Q})$ and $\mathrm{D}_f^{Rp}(\mathbb{P}, \mathbb{Q}) \leq \mathrm{D}_f^{Ra}(\mathbb{P}, \mathbb{Q})$

Proof. Showing that $\mathrm{D}^S(\mathbb{P}, \mathbb{Q}) \leq \mathrm{D}_f^{Rp}(\mathbb{P}, \mathbb{Q})$:
Let

$$C_S^*(x) = \arg\sup_{C:\mathcal{X}\to\mathbb{R}} \mathbb{E}_{x\sim\mathbb{P}}\left[f(C(x))\right] + \mathbb{E}_{z\sim\mathbb{Q}}\left[f(-C(y))\right]$$

and

$$C_{Rp}^*(x) = \arg\sup_{C:\mathcal{X}\to\mathbb{R}} \mathop{\mathbb{E}}_{\substack{x\sim\mathbb{P} \\ y\sim\mathbb{Q}}}\left[f\left(C(x) - C(y)\right)\right].$$

$$
\begin{aligned}
\mathrm{D}^S(\mathbb{P},\mathbb{Q}) &= \sup_{C:\mathcal{X}\to\mathbb{R}} \mathbb{E}_{x\sim\mathbb{P}}\left[f(C(x))\right] + \mathbb{E}_{z\sim\mathbb{Q}}\left[f(-C(y))\right] \\
&= 2 \underset{\substack{x\sim\mathbb{P}\\ y\sim\mathbb{Q}}}{\mathbb{E}}\left[\frac{1}{2} f\left(C^*_S(x)\right) + \frac{1}{2} f\left(-C^*_S(y)\right)\right] \\
&\leq 2 \underset{\substack{x\sim\mathbb{P}\\ y\sim\mathbb{Q}}}{\mathbb{E}}\left[f\left(\frac{1}{2}C^*_S(x) - \frac{1}{2}C^*_S(y)\right)\right] \\
&= 2 \underset{\substack{x\sim\mathbb{P}\\ y\sim\mathbb{Q}}}{\mathbb{E}}\left[f\left(C'(x) - C'(y)\right)\right], \text{ where } C'(x) = \frac{1}{2}C^*_S(x) \\
&\leq \sup_{C:\mathcal{X}\to\mathbb{R}} 2 \underset{\substack{x\sim\mathbb{P}\\ y\sim\mathbb{Q}}}{\mathbb{E}}\left[f\left(C(x) - C(y)\right)\right] \\
&= \mathrm{D}^{Rp}_f(\mathbb{P},\mathbb{Q})
\end{aligned}
$$

Showing that $\mathrm{D}^{Rp}_f(\mathbb{P},\mathbb{Q}) \leq \mathrm{D}^{Ralf}_f(\mathbb{P},\mathbb{Q})$:

$$
\begin{aligned}
\mathrm{D}^{Rp}_f(\mathbb{P},\mathbb{Q}) &= \underset{C:\mathcal{X}\to\mathbb{R}}{\arg\sup}\, 2 \underset{\substack{x\sim\mathbb{P}\\ y\sim\mathbb{Q}}}{\mathbb{E}}\left[f\left(C(x) - C(y)\right)\right] \\
&= 2 \underset{\substack{x\sim\mathbb{P}\\ y\sim\mathbb{Q}}}{\mathbb{E}}\left[f\left(C^*_{Rp}(x) - C^*_{Rp}(y)\right)\right] \\
&= 2 \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[\underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[f\left(C^*_{Rp}(x) - C^*_{Rp}(y)\right)|x\right]\right] \\
&\leq 2 \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(\underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[C^*_{Rp}(x) - C^*_{Rp}(y)|x\right]\right)\right] \\
&= 2 \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(C^*_{Rp}(x) - \underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[C^*_{Rp}(y)\right]\right)\right] \\
&\leq \sup_{C:\mathcal{X}\to\mathbb{R}} 2 \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(C(x) - \underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[C(y)\right]\right)\right] \\
&= \mathrm{D}^{Ralf}_f(\mathbb{P},\mathbb{Q})
\end{aligned}
$$

Showing that $\mathrm{D}^{Rp}_f(\mathbb{P},\mathbb{Q}) \leq \mathrm{D}^{Ra}_f(\mathbb{P},\mathbb{Q})$:

$$
\begin{aligned}
\mathrm{D}^{Rp}_f(\mathbb{P},\mathbb{Q}) &= \underset{C:\mathcal{X}\to\mathbb{R}}{\arg\sup}\, 2 \underset{\substack{x\sim\mathbb{P}\\ y\sim\mathbb{Q}}}{\mathbb{E}}\left[f\left(C(x) - C(y)\right)\right] \\
&= 2 \underset{\substack{x\sim\mathbb{P}\\ y\sim\mathbb{Q}}}{\mathbb{E}}\left[f\left(C^*_{Rp}(x) - C^*_{Rp}(y)\right)\right] \\
&= \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[\underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[f\left(C^*_{Rp}(x) - C^*_{Rp}(y)\right)|x\right]\right] + \underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[\underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(C^*_{Rp}(x) - C^*_{Rp}(y)\right)|y\right]\right] \\
&\leq \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(\underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[C^*_{Rp}(x) - C^*_{Rp}(y)|x\right]\right)\right] + \underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[f\left(\underset{x\sim\mathbb{P}}{\mathbb{E}}\left[C^*_{Rp}(x) - C^*_{Rp}(y)|y\right]\right)\right] \\
&= \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(C^*_{Rp}(x) - \underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[C^*_{Rp}(y)\right]\right)\right] + \underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[f\left(\underset{x\sim\mathbb{P}}{\mathbb{E}}\left[C^*_{Rp}(x)\right] - C^*_{Rp}(y)\right)\right] \\
&\leq \sup_{C:\mathcal{X}\to\mathbb{R}} \underset{x\sim\mathbb{P}}{\mathbb{E}}\left[f\left(C(x) - \underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[C(y)\right]\right)\right] + \underset{y\sim\mathbb{Q}}{\mathbb{E}}\left[f\left(\underset{x\sim\mathbb{P}}{\mathbb{E}}\left[C(x)\right] - C(y)\right)\right] \\
&= \mathrm{D}^{Ra}_f(\mathbb{P},\mathbb{Q})
\end{aligned}
$$

□

C. Bias in RalfGANs, RaGANs, and RcGANs

Note that we refer to the second term in RaGANs as "RaGAN2". When possible, we calculate the bias for RalfGANs, RaGAN2s, RaGANs, and RcGANs.

Let
$\underset{x\sim\mathbb{P}}{\mathbb{E}}[C(x)] = \mu_x$,
$\underset{x\sim\mathbb{P}}{Var}[C(x)] = \sigma_x^2$,
$\underset{x\sim\mathbb{P}}{\mathbb{E}}[C(x)^2] = \sigma_x^2 + \mu_x^2$,

$\underset{y\sim\mathbb{Q}}{\mathbb{E}}[C(y)] = \mu_y$,
$\underset{y\sim\mathbb{Q}}{Var}[C(y)] = \sigma_y^2$,
$\underset{y\sim\mathbb{Q}}{\mathbb{E}}[C(y)^2] = \sigma_y^2 + \mu_y^2$.

In a minibatch of size k, we have that $x_1, \ldots, x_k$ and $y_1, \ldots, y_k$ are iid.
Thus, $C(x_1), \ldots, C(x_k)$ and $C(y_1), \ldots, C(y_k)$ are also iid.
This means that:
$\mathbb{E}[C(x_i)C(x_j)] = \mathbb{E}[C(x_i)]\mathbb{E}[C(x_j)] = \mu_x^2 \ \forall i \neq j$,
$\mathbb{E}[C(y_i)C(y_j)] = \mathbb{E}[C(y_i)]\mathbb{E}[C(y_j)] = \mu_y^2 \ \forall i \neq j$.

C.1. SGAN

$$f(x) = \log(\operatorname{sigmoid}(x)) + \log(2) = -\log(1+e^{-x}) + \log(2)$$

$$\begin{aligned}
\text{Bias}^{RaSGAN}(\mathbb{P},\mathbb{Q}) &= \mathbb{E}\left[f\left(C(x) - \frac{1}{k}\sum_{i=1}^{k} C(y_i)\right) - f\left(C(x)-\mu_y\right)\right] \\
&= \mathbb{E}\left[-\log\left(1+e^{\frac{1}{k}\sum_{i=1}^{k} C(y_i) - C(x)}\right) + \log(2) + \log\left(1+e^{\mu_y - C(x)}\right) - \log(2)\right] \\
&= \mathbb{E}\left[\log\left(\frac{1+e^{\mu_y - C(x)}}{1+e^{\frac{1}{k}\sum_{i=1}^{k} C(y_i) - C(x)}}\right)\right] \\
&= \mathbb{E}\left[\log\left(\frac{e^{C(x)}+e^{\mu_y}}{e^{C(x)}+e^{\frac{1}{k}\sum_{i=1}^{k} C(y_i)}}\right)\right] \\
&= \mathbb{E}\left[\log\left(e^{C(x)}+e^{\mu_y}\right) - \log\left(e^{C(x)}+e^{\frac{1}{k}\sum_{i=1}^{k} C(y_i)}\right)\right] \\
&\approx \mathbb{E}\left[C(x) + e^{\mu_y - C(x)} - C(x) - e^{\frac{1}{k}\sum_{i=1}^{k} C(y_i) - C(x)}\right] \\
&= \mathbb{E}\left[\frac{e^{\mu_y} - e^{\frac{1}{k}\sum_{i=1}^{k} C(y_i)}}{e^{C(x)}}\right]
\end{aligned}$$

We cannot find a close form for the bias.

C.2. LSGAN

$$f(x) = -(x-1)^2 + 1$$

$$
\begin{aligned}
\widehat{\mathrm{Div}}^{\mathrm{RaLSGAN}}(\mathbb{P},\mathbb{Q}) &= \mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k} f\left(C(x_i)-\frac{1}{k}\sum_{j=1}^{k}C(y_j)\right)\right] \\
&= \mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k}\left(-\left(C(x_i)-\frac{1}{k}\sum_{j=1}^{k}C(y_j)-1\right)^2+1\right)\right] \\
&= \mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k}\left(-C(x_i)^2+\frac{2}{k}\sum_{j=1}^{k}C(x_i)C(y_j)+2C(x_i)-2\frac{1}{k}\sum_{j=1}^{k}C(y_j)-\frac{1}{k^2}\left(\sum_{j=1}^{k}C(y_j)\right)^2\right)\right] \\
&= \frac{1}{k}\sum_{i=1}^{k}\left(-\mathbb{E}\left[C(x_i)^2\right]+\frac{2}{k}\sum_{j=1}^{k}\mathbb{E}\left[C(x_i)\right]\mathbb{E}\left[C(y_j)\right]+2\mathbb{E}\left[C(x_i)\right]-2\frac{1}{k}\sum_{j=1}^{k}\mathbb{E}\left[C(y_j)\right]\right. \\
&\left.-\frac{1}{k^2}\sum_{j=1}^{k}\mathbb{E}[C(y_j)^2]-\frac{1}{k^2}\sum_{\substack{r=1\\ r\neq j}}^{k}\sum_{j=1}^{k}\mathbb{E}[C(y_j)]\mathbb{E}[C(y_r)]\right) \\
&= \frac{1}{k}\sum_{i=1}^{k}\left(-\sigma_x^2-\mu_x^2+2\mu_x\mu_y+2\mu_x-2\mu_y-\frac{1}{k}(\sigma_y^2+\mu_y^2)-\frac{(k-1)}{k}\mu_y^2\right) \\
&= -\sigma_x^2-\mu_x^2+2\mu_x\mu_y+2\mu_x-2\mu_y-\frac{1}{k}\sigma_y^2-\mu_y^2
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\mathrm{Div}}^{\mathrm{RaLSGAN2}}(\mathbb{P},\mathbb{Q}) &= \mathbb{E}\left[\frac{1}{k}\sum_{j=1}^{k} f\left(\frac{1}{k}\sum_{i=1}^{k}C(x_i)-C(y_j)\right)\right] \\
&= \mathbb{E}\left[\frac{1}{k}\sum_{j=1}^{k}\left(-\left(\frac{1}{k}\sum_{i=1}^{k}C(x_i)-C(y_j)-1\right)^2-1\right)\right] \\
&= \mathbb{E}\left[\frac{1}{k}\sum_{j=1}^{k}\left(-C(y_j)^2+\frac{2}{k}\sum_{x=1}^{k}C(x_i)C(y_j)-2C(y_j)+2\frac{1}{k}\sum_{i=1}^{k}C(x_i)-\frac{1}{k^2}\left(\sum_{i=1}^{k}C(x_i)\right)^2\right)\right] \\
&= \frac{1}{k}\sum_{j=1}^{k}\left(-\mathbb{E}\left[C(y_j)^2\right]+\frac{2}{k}\sum_{i=1}^{k}\mathbb{E}\left[C(x_i)\right]\mathbb{E}\left[C(y_j)\right]-2\mathbb{E}\left[C(y_j)\right]+2\frac{1}{k}\sum_{i=1}^{k}\mathbb{E}\left[C(x_i)\right]\right. \\
&\left.-\frac{1}{k^2}\sum_{i=1}^{k}\mathbb{E}[C(x_i)^2]-\frac{1}{k^2}\sum_{\substack{r=1\\ r\neq i}}^{k}\sum_{i=1}^{k}\mathbb{E}[C(x_i)]\mathbb{E}[C(x_r)]\right) \\
&= \frac{1}{k}\sum_{j=1}^{k}\left(-\sigma_y^2-\mu_y^2+2\mu_x\mu_y-2\mu_y+2\mu_x-\frac{1}{k}(\sigma_x^2+\mu_x^2)-\frac{(k-1)}{k}\mu_x^2\right) \\
&= -\sigma_y^2-\mu_y^2+2\mu_x\mu_y-2\mu_y+2\mu_x-\frac{1}{k}\sigma_x^2-\mu_x^2
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Div}^{\mathrm{RaLSGAN}}(\mathbb{P},\mathbb{Q}) &= \mathbb{E}\left[f\left(C(x)-\mu_y\right)\right] \\
&= \mathbb{E}\left[-\left(C(x)-\mu_y-1\right)^2-1\right] \\
&= \mathbb{E}\left[-C(x)^2+2C(x)\mu_y+2C(x)-2\mu_y-\mu_y^2\right] \\
&= -\sigma_x^2-\mu_x^2+2\mu_x\mu_y+2\mu_x-2\mu_y-\mu_y^2
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Div}^{\mathrm{RaLSGAN2}}(\mathbb{P},\mathbb{Q}) &= \mathbb{E}\left[f\left(\mu_x-C(y)\right)\right] \\
&= \mathbb{E}\left[-\left(\mu_x-C(y)-1\right)^2-1\right] \\
&= \mathbb{E}\left[-\mu_x^2+2C(y)\mu_x-2C(y)+2\mu_x-C(y)^2\right] \\
&= -\sigma_y^2-\mu_y^2+2\mu_x\mu_y-2\mu_y+2\mu_x-\mu_x^2
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Bias}^{\mathrm{RaLSGAN}}(\mathbb{P},\mathbb{Q}) &= \widehat{\mathrm{Div}}^{\mathrm{RaLSGAN}}(\mathbb{P},\mathbb{Q})-\mathrm{Div}^{\mathrm{RaLSGAN}}(\mathbb{P},\mathbb{Q}) \\
&= -\sigma_x^2-\mu_x^2+2\mu_x\mu_y+2\mu_x-2\mu_y-\frac{1}{k}\sigma_y^2-\mu_y^2+\sigma_x^2+\mu_x^2-2\mu_x\mu_y-2\mu_x+2\mu_y+\mu_y^2 \\
&= -\frac{1}{k}\sigma_y^2
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Bias}^{\mathrm{RaLSGAN2}}(\mathbb{P},\mathbb{Q}) &= \widehat{\mathrm{Div}}^{\mathrm{RaLSGAN2}}(\mathbb{P},\mathbb{Q})-\mathrm{Div}^{\mathrm{RaLSGAN2}}(\mathbb{P},\mathbb{Q}) \\
&= -\sigma_y^2-\mu_y^2+2\mu_x\mu_y-2\mu_y+2\mu_x-\frac{1}{k}\sigma_x^2-\mu_x^2+\sigma_y^2+\mu_y^2-2\mu_x\mu_y+2\mu_y-2\mu_x+\mu_x^2 \\
&= -\frac{1}{k}\sigma_x^2
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{Bias}^{RalfLSGAN} &= \mathrm{Bias}^{RaLSGAN}(\mathbb{P},\mathbb{Q})+\mathrm{Bias}^{RaLSGAN2}(\mathbb{Q},\mathbb{P}) \\
&= -\frac{1}{k}\sigma_y^2-\frac{1}{k}\sigma_x^2 \\
&= -\frac{1}{k}\left(\sigma_x^2+\sigma_y^2\right)
\end{aligned}
$$

Let
$\hat{\sigma}_x^2=\frac{1}{(k-1)}\sum_{i=1}^k\left(C(x_i)-\frac{1}{k}\sum_{i=1}^k C(x_j)\right)$,
$\hat{\sigma}_y^2=\frac{1}{(k-1)}\sum_{i=1}^k\left(C(y_i)-\frac{1}{k}\sum_{i=1}^k C(y_j)\right)$.

We know that $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$ are unbiased estimators of σ_x^2 and σ_y^2 respectively.
Thus, if we add $\frac{1}{k}\hat{\sigma}_y^2$ to the objective function of RalfLSGAN and $\frac{1}{k}(\hat{\sigma}_x^2+\hat{\sigma}_y^2)$ to the objective function of RaLSGAN, we have that the new objective functions are unbiased.

$$
\begin{aligned}
\widehat{\mathrm{Div}}^{\mathrm{RcLSGAN}}(\mathbb{P},\mathbb{Q}) &= \mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k} f\left(C(x_i) - \frac{1}{2k}\sum_{j=1}^{k}\left(C(x_j)+C(y_j)\right)\right)\right] \\
&= \mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k}\left(-\left(C(x_i) - \frac{1}{2k}\sum_{j=1}^{k}\left(C(x_j)+C(y_j)\right)-1\right)^2+1\right)\right] \\
&= \mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k}\left(-C(x_i)^2 + \frac{1}{k}\sum_{j=1}^{k} C(x_i)\left(C(x_j)+C(y_j)\right) + 2C(x_i) - \frac{1}{k}\sum_{j=1}^{k} C(x_j) - \frac{1}{k}\sum_{j=1}^{k} C(y_j)\right.\right. \\
&\left.\left. - \frac{1}{4k^2}\left(\sum_{j=1}^{k} C(x_j)+C(y_j)\right)^2\right)\right] \\
&= \frac{1}{k}\sum_{i=1}^{k}\left(-\mathbb{E}\left[C(x_i)^2\right] + \frac{1}{k}\mathbb{E}\left[C(x_i)^2\right] + \frac{1}{k}\sum_{\substack{j=1\\ j\neq i}}^{k}\mathbb{E}\left[C(x_i)\right]\mathbb{E}\left[C(x_j)\right] + \frac{1}{k}\sum_{j=1}^{k}\mathbb{E}\left[C(x_i)\right]\mathbb{E}\left[C(y_j)\right]\right. \\
& +2\mathbb{E}\left[C(x_i)\right] - \frac{1}{k}\sum_{j=1}^{k}\mathbb{E}\left[C(x_j)\right] - \frac{1}{k}\sum_{j=1}^{k}\mathbb{E}\left[C(y_j)\right] - \frac{1}{4k^2}\sum_{j=1}^{k}\mathbb{E}[(C(x_j)+C(y_j))^2] \\
&\left. - \frac{1}{4k^2}\sum_{\substack{r=1\\ r\neq j}}^{k}\sum_{j=1}^{k}\mathbb{E}[C(x_i)+C(y_i)]\mathbb{E}[C(x_r)+C(y_r)]\right) \\
&= \left(\frac{1}{k}-1\right)\left(\sigma_x^2+\mu_x^2\right) + \frac{(k-1)}{k}\mu_x^2 + \mu_x\mu_y + 2\mu_x - \mu_x - \mu_y \\
& - \frac{1}{4k}((\sigma_x^2+\mu_x^2) + 2\mu_x\mu_y + (\sigma_y^2+\mu_y^2)) - \frac{(k-1)}{4k}(\mu_x^2 + 2\mu_x\mu_y + \mu_y^2) \\
&= \frac{(1-k)}{k}\sigma_x^2 + \mu_x\mu_y + \mu_x - \mu_y - \frac{1}{4}\mu_x^2 - \frac{1}{2}\mu_x\mu_y - \frac{1}{4}\mu_y^2 - \frac{1}{4k}\sigma_x^2 - \frac{1}{4k}\sigma_y^2 \\
&= \frac{(.75-k)}{k}\sigma_x^2 - \frac{1}{4k}\sigma_y^2 - \frac{1}{4}\mu_x^2 - \frac{1}{4}\mu_y^2 + \frac{1}{2}\mu_x\mu_y + \mu_x - \mu_y
\end{aligned}
$$

$$
\begin{aligned}
\widehat{\text{Div}}^{\text{RcLSGAN}}(\mathbb{P},\mathbb{Q}) &= \mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k}\left(-\left(C(y_i)-\frac{1}{2k}\sum_{j=1}^{k}(C(x_j)+C(y_j))+1\right)^2+1\right)\right] \\
&= \mathbb{E}\left[\frac{1}{k}\sum_{i=1}^{k}\left(-C(y_i)^2+\frac{1}{k}\sum_{j=1}^{k}C(y_i)\left(C(x_j)+C(y_j)\right)-2C(y_i)+\frac{1}{k}\sum_{j=1}^{k}C(x_j)+\frac{1}{k}\sum_{j=1}^{k}C(y_j)\right.\right. \\
&\left.\left.-\frac{1}{4k^2}\left(\sum_{j=1}^{k}C(x_j)+C(y_j)\right)^2\right)\right] \\
&= \frac{1}{k}\sum_{i=1}^{k}\left(-\mathbb{E}\left[C(y_i)^2\right]+\frac{1}{k}\mathbb{E}\left[C(y_i)^2\right]+\frac{1}{k}\sum_{\substack{j=1\\j\neq i}}^{k}\mathbb{E}\left[C(y_i)\right]\mathbb{E}\left[C(y_j)\right]+\frac{1}{k}\sum_{j=1}^{k}\mathbb{E}\left[C(x_i)\right]\mathbb{E}\left[C(y_j)\right]\right. \\
&-2\mathbb{E}\left[C(y_i)\right]+\frac{1}{k}\sum_{j=1}^{k}\mathbb{E}\left[C(x_j)\right]+\frac{1}{k}\sum_{j=1}^{k}\mathbb{E}\left[C(y_j)\right]-\frac{1}{4k^2}\sum_{j=1}^{k}\mathbb{E}[(C(x_j)+C(y_j))^2] \\
&\left.-\frac{1}{4k^2}\sum_{\substack{r=1\\r\neq j}}^{k}\sum_{j=1}^{k}\mathbb{E}[C(x_i)+C(y_i)]\mathbb{E}[C(x_r)+C(y_r)]\right) \\
&= \left(\frac{1}{k}-1\right)\left(\sigma_y^2+\mu_y^2\right)+\frac{(k-1)}{k}\mu_y^2+\mu_x\mu_y-2\mu_y+\mu_x+\mu_y \\
&-\frac{1}{4k}((\sigma_x^2+\mu_x^2)+2\mu_x\mu_y+(\sigma_y^2+\mu_y^2))-\frac{(k-1)}{4k}(\mu_x^2+2\mu_x\mu_y+\mu_y^2) \\
&= \frac{(1-k)}{k}\sigma_y^2+\mu_x\mu_y+\mu_x-\mu_y-\frac{1}{4}\mu_x^2-\frac{1}{2}\mu_x\mu_y-\frac{1}{4}\mu_y^2-\frac{1}{4k}\sigma_x^2-\frac{1}{4k}\sigma_y^2 \\
&= \frac{(.75-k)}{k}\sigma_y^2-\frac{1}{4k}\sigma_x^2-\frac{1}{4}\mu_x^2-\frac{1}{4}\mu_y^2+\frac{1}{2}\mu_x\mu_y+\mu_x-\mu_y
\end{aligned}
$$

$$
\begin{aligned}
\text{Div}^{\text{RcLSGAN}}(\mathbb{P},\mathbb{Q}) &= \mathbb{E}\left[f\left(C(x)-\frac{(\mu_x+\mu_y)}{2}\right)\right] \\
&= \mathbb{E}\left[-\left(C(x)-\frac{(\mu_x+\mu_y)}{2}-1\right)^2-1\right] \\
&= \mathbb{E}\left[-C(x)^2+C(x)(\mu_x+\mu_y)+2C(x)-(\mu_x+\mu_y)-\frac{(\mu_x+\mu_y)^2}{4}\right] \\
&= -\sigma_x^2-\mu_x^2+\mu_x^2+\mu_x\mu_y+\mu_x-\mu_y-\frac{1}{4}(\mu_x^2+\mu_y^2+2\mu_x\mu_y) \\
&= -\sigma_x^2+\frac{1}{2}\mu_x\mu_y+\mu_x-\mu_y-\frac{1}{4}\mu_x^2-\frac{1}{4}\mu_y^2
\end{aligned}
$$

$$\begin{aligned}
\text{Div}^{\text{RcLSGAN2}}(\mathbb{P},\mathbb{Q}) &= \mathbb{E}\left[f\left(C(x)-\mu_y\right)\right] \\
&= \mathbb{E}\left[-\left(C(y)-\frac{(\mu_x+\mu_y)}{2}+1\right)^2-1\right] \\
&= \mathbb{E}\left[-C(y)^2+C(y)(\mu_x+\mu_y)-2C(y)+(\mu_x+\mu_y)-\frac{(\mu_x+\mu_y)^2}{4}\right] \\
&= -\sigma_y^2-\mu_y^2+\mu_y^2+\mu_x\mu_y+\mu_x-\mu_y-\frac{1}{4}(\mu_x^2+\mu_y^2+2\mu_x\mu_y) \\
&= -\sigma_x^2+\frac{1}{2}\mu_x\mu_y+\mu_x-\mu_y-\frac{1}{4}\mu_x^2-\frac{1}{4}\mu_y^2
\end{aligned}$$

$$\begin{aligned}
\text{Bias}^{\text{RaLSGAN}}(\mathbb{P},\mathbb{Q}) &= \widehat{\text{Div}}^{\text{RaLSGAN}}(\mathbb{P},\mathbb{Q})-\text{Div}^{\text{RaLSGAN}}(\mathbb{P},\mathbb{Q}) \\
&= \frac{3}{4k}\sigma_x^2-\frac{1}{4k}\sigma_y^2
\end{aligned}$$

$$\begin{aligned}
\text{Bias}^{\text{RaLSGAN2}}(\mathbb{P},\mathbb{Q}) &= \widehat{\text{Div}}^{\text{RaLSGAN2}}(\mathbb{P},\mathbb{Q})-\text{Div}^{\text{RaLSGAN2}}(\mathbb{P},\mathbb{Q}) \\
&= \frac{3}{4k}\sigma_y^2-\frac{1}{4k}\sigma_x^2
\end{aligned}$$

$$\begin{aligned}
\text{Bias}^{RalfLSGAN} &= \text{Bias}^{RaLSGAN}(\mathbb{P},\mathbb{Q})+\text{Bias}^{RaLSGAN2}(\mathbb{Q},\mathbb{P}) \\
&= \frac{3}{4k}\sigma_x^2-\frac{1}{4k}\sigma_y^2+\frac{3}{4k}\sigma_y^2-\frac{1}{4k}\sigma_x^2 \\
&= \frac{1}{2k}\left(\sigma_x^2+\sigma_y^2\right)
\end{aligned}$$

Let
$\hat{\sigma}_x^2 = \frac{1}{(k-1)}\sum_{i=1}^k \left(C(x_i)-\frac{1}{k}\sum_{i=1}^k C(x_j)\right)$,
$\hat{\sigma}_y^2 = \frac{1}{(k-1)}\sum_{i=1}^k \left(C(y_i)-\frac{1}{k}\sum_{i=1}^k C(y_j)\right)$.

We know that $\hat{\sigma}_x^2$ and $\hat{\sigma}_y^2$ are unbiased estimators of σ_x^2 and σ_y^2 respectively.
Thus, if we subtract $\frac{1}{2k}(\hat{\sigma}_x^2+\hat{\sigma}_y^2)$ to the objective function of RcLSGAN, we have that the new objective functions are unbiased.

C.3. HINGEGAN

$$f(x) = -\max(0, 1-x)+1$$

For simplicity:
Let $x' = C(x)$, $y'_i = C(y_i)$, $p(x)$ and $q(x)$ be the probability density functions of x' and y'_i.

$$\begin{aligned}\mathrm{Div}^{\mathrm{RaHingeGAN}}(\mathbb{P},\mathbb{Q}) &= \mathbb{E}\left[f\left(C(x)-\frac{1}{k}\sum_{i=1}^{k}C(y_i)\right)\right]\\ &= \mathbb{E}\left[-\max\left(0,1+\frac{1}{k}\sum_{i=1}^{k}y_i'-x'\right)+1\right]\\ &= -\int_{-\infty}^{\infty}...\int_{-\infty}^{\infty}\int_{-\infty}^{1+\frac{1}{k}\sum_{i=1}^{k}y_i'}\left(1+\frac{1}{k}\sum_{i=1}^{k}y_i'-x\right)p(x)q(y)...q(y)dxdy_1...dy_k\end{aligned}$$

This is non-linear and we cannot derive a close-form.

D. Architecture

Generator
$z \in \mathbb{R}^{128} \sim N(0, I)$
linear, 128 -¿ 512*4*4
Reshape, 512*4*4 -¿ 512 x 4 x 4
ConvTranspose2d 4x4, stride 2, pad 1, 512-¿256
BN and ReLU
ConvTranspose2d 4x4, stride 2, pad 1, 256-¿128
BN and ReLU
ConvTranspose2d 4x4, stride 2, pad 1, 128-¿64
BN and ReLU
ConvTranspose2d 3x3, stride 1, pad 1, 64-¿3
Tanh

Discriminator
$x \in \mathbb{R}^{3x32x32}$
Conv2d 3x3, stride 1, pad 1, 3-¿64
LeakyReLU 0.1
Conv2d 4x4, stride 2, pad 1, 64-¿64
LeakyReLU 0.1
Conv2d 3x3, stride 1, pad 1, 64-¿128
LeakyReLU 0.1
Conv2d 4x4, stride 2, pad 1, 128-¿128
LeakyReLU 0.1
Conv2d 3x3, stride 1, pad 1, 128-¿256
LeakyReLU 0.1
Conv2d 4x4, stride 2, pad 1, 256-¿256
LeakyReLU 0.1
Conv2d 3x3, stride 1, pad 1, 256-¿512
Reshape, 512 x 4 x 4 -¿ 512*4*4
linear, 512*4*4 -¿ 1