Appendix: Lipschitz Continuity in Model-based Reinforcement Learning

We first restate the core lemmas, theorems, and claims presented in our paper below:

Lemma 1. *A generalized transition function $\widehat{T}_{\mathcal{G}}$ induced by a Lipschitz model class F_g is Lipschitz with a constant:*

$$K^{\mathcal{A}}_{W,W}(\widehat{T}_{\mathcal{G}}) := \sup_{a} \sup_{\mu_1,\mu_2} \frac{W\big(\widehat{T}_{\mathcal{G}}(\cdot|\mu_1,a),\widehat{T}_{\mathcal{G}}(\cdot|\mu_2,a)\big)}{W(\mu_1,\mu_2)} \leq K_F$$

Lemma 2. *(Composition Lemma) Define three metric spaces (M_1,d_1), (M_2,d_2), and (M_3,d_3). Define Lipschitz functions $f: M_2 \mapsto M_3$ and $g: M_1 \mapsto M_2$ with constants $K_{d_2,d_3}(f)$ and $K_{d_1,d_2}(g)$. Then, $h: f \circ g: M_1 \mapsto M_3$ is Lipschitz with constant $K_{d_1,d_3}(h) \leq K_{d_2,d_3}(f)K_{d_1,d_2}(g)$.*

Theorem 1. *Define a Δ-accurate $\widehat{T}_{\mathcal{G}}$ with the Lipschitz constant K_F and an MDP with a Lipschitz transition function $T_{\mathcal{G}}$ with constant K_T. Let $\bar{K} = \min\{K_F, K_T\}$. Then $\forall n \geq 1$:*

$$\delta(n) := W\big(\widehat{T}^n_{\mathcal{G}}(\cdot \mid \mu), T^n_{\mathcal{G}}(\cdot \mid \mu)\big) \leq \Delta \sum_{i=0}^{n-1} (\bar{K})^i \ .$$

Theorem 2. *Assume a Lipschitz model class F_g with a Δ-accurate $\widehat{T}$ with $\bar{K} = \min\{K_F, K_T\}$. Further, assume a Lipschitz reward function with constant $K_R = K_{d_{\mathcal{S}},\mathbb{R}}(R)$. Then $\forall s \in \mathcal{S}$ and $\bar{K} \in [0, \frac{1}{\gamma})$*

$$\big|V_T(s) - V_{\widehat{T}}(s)\big| \leq \frac{\gamma K_R \Delta}{(1-\gamma)(1-\gamma\bar{K})} \ .$$

Lemma 3. *Given a Lipschitz function $f: \mathcal{S} \mapsto \mathbb{R}$ with constant $K_{d_{\mathcal{S}},d_{\mathbb{R}}}(f)$:*

$$K^{\mathcal{A}}_{d_{\mathcal{S}},d_{\mathbb{R}}}\Big(\int \widehat{T}(s'|s,a)f(s')ds'\Big) \leq K_{d_{\mathcal{S}},d_{\mathbb{R}}}(f) K^{\mathcal{A}}_{d_{\mathcal{S}},W}\big(\widehat{T}\big) \ .$$

Lemma 4. *The following operators (Asadi & Littman, 2017) are Lipschitz with constants:*

1. $K_{\|\|_\infty,d_R}(\max(x)) = K_{\|\|_\infty,d_R}\big(mean(x)\big) = K_{\|\|_\infty,d_R}(\epsilon\text{-}greedy(x)) = 1$
2. $K_{\|\|_\infty,d_R}(mm_\beta(x) := \frac{\log \frac{\sum_i e^{\beta x_i}}{n}}{\beta}) = 1$
3. $K_{\|\|_\infty,d_R}(boltz_\beta(x) := \frac{\sum_{i=1}^n x_i e^{\beta x_i}}{\sum_{i=1}^n e^{\beta x_i}}) \leq \sqrt{|A|} + \beta V_{\max}|A|$

Theorem 3. *For any non-expansion backup operator f outlined in Lemma 4, GVI computes a value function with a Lipschitz constant bounded by $\frac{K^{\mathcal{A}}_{d_{\mathcal{S}},d_R}(R)}{1-\gamma K_{d_{\mathcal{S}},W}(T)}$ if $\gamma K^{\mathcal{A}}_{d_{\mathcal{S}},W}(T) < 1$.*

We now provide proofs of various results mentioned in the paper:

Claim 1. *In a finite MDP, transition probabilities can be expressed using a finite set of deterministic functions and a distribution over the functions.*

Proof. Let $Pr(s,a,s')$ denote the probability of a transiton from s to s' when executing the action a. Define an ordering over states $s_1, ..., s_n$ with an additional unreachable state s_0. Now define the cumulative probability distribution:

$$C(s,a,s_i) := \sum_{j=0}^{i} Pr(s,a,s_j) \ .$$

Further define L as the set of distinct entries in C:

$$L := \Big\{ C(s,a,s_i) \big| \quad s \in \mathcal{S}, i \in [0,n] \Big\} \ .$$

Note that, since the MDP is assumed to be finite, then $|L|$ is finite. We sort the values of L and denote, by c_i, ith smallest value of the set. Note that $c_0 = 0$ and $c_{|L|} = 1$. We now build determinstic set of functions $f_1, ..., f_{|L|}$ as follows: $\forall i = 1$ to $|L|$ and $\forall j = 1$ to n, define $f_i(s) = s_j$ if and only if:

$$C(s, a, s_{j-1}) < c_i \leq C(s, a, s_j) \ .$$

We also define the probability distribution g over f as follows:

$$g(f_i|a) := c_i - c_{i-1} \ .$$

Given the functions $f_1, ..., f_{|L|}$ and the distribution g, we can now compute the probability of a transition to s_j from s after executing action a:

$$\begin{aligned}
&\sum_i \mathbb{1}(f_i(s) = s_j) \ g(f_i|a) \\
&= \sum_i \mathbb{1}\big(C(s, a, s_{j-1}) < c_i \leq C(s, a, s_j)\big) \ (c_i - c_{i-1}) \\
&= C(s, a, s_j) - C(s, a, s_{j-1}) \\
&= Pr(s, a, s_j) \ ,
\end{aligned}$$

where $\mathbb{1}$ is a binary function that outputs one if and only if its condition holds. We reconstructed the transition probabilities using distribution g and deterministic functions $f_1, ..., f_{|L|}$. $\square$

Claim 2. *Given a deterministic and linear transition model, and a linear reward signal, the bounds provided in Theorems 1 and 2 are both tight.*

Assume a linear transition function T defined as:

$$T(s) = Ks$$

Assume our learned transition function $\hat{T}$:

$$\hat{T}(s) := Ks + \Delta$$

Note that:

$$\max_s \big|T(s) - \hat{T}(s)\big| = \Delta$$

and that:

$$\min\{K_T, K_{\hat{T}}\} = K$$

First observe that the bound in Theorem 2 is tight for $n = 2$:

$$\forall s \quad \Big|T\big(T(s)\big) - \hat{T}\big(\hat{T}(s)\big)\Big| = \Big|K^2 s - K^2 s + \Delta(1 + K)\Big| = \Delta \sum_{i=0}^{1} K^i$$

and more generally and after n compositions of the models, denoted by T^n and $\hat{T}^n$, the following equality holds:

$$\forall s \quad \Big|T^n(s) - \hat{T}^n(s)\Big| = \Delta \sum_{i=0}^{n-1} K^i$$

Lets further assume that the reward is linear:

$$R(s) = K_R s$$

Consider the state $s = 0$. Note that clearly $v(0) = 0$. We now compute the value predicted using $\hat{T}$, denoted by $\hat{v}(0)$:

$$\begin{aligned}
\hat{v}(0) &= R(0) + \gamma R(0 + \Delta \sum_{i=0}^{0} K^i) + \gamma^2 R(0 + \Delta \sum_{i=0}^{1} K^i) + \gamma^3 R(0 + \Delta \sum_{i=0}^{2} K^i) + ... \\
&= 0 + \gamma K_R \Delta \sum_{i=0}^{0} K^i + \gamma^2 K_R \Delta \sum_{i=0}^{1} K^i) + \gamma^3 K_R \Delta \sum_{i=0}^{2} K^i + ...
\end{aligned}$$

$$= \quad \gamma K_R \Delta \sum_{n=0}^{\infty} \gamma^n \sum_{i=0}^{n-1} K^i = \frac{\gamma K_R \Delta}{(1-\gamma)(1-\gamma \bar{K})} \ ,$$

and so:

$$|v(0) - \hat{v}(0)| = \frac{\gamma K_R \Delta}{(1-\gamma)(1-\gamma \bar{K})}$$

Note that this exactly matches the bound derived in our Theorem 2.

Lemma 1. *A generalized transition function $\widehat{T}_{\mathcal{G}}$ induced by a Lipschitz model class F_g is Lipschitz with a constant:*

$$K^{\mathcal{A}}_{W,W}(\widehat{T}_{\mathcal{G}}) := \sup_a \sup_{\mu_1,\mu_2} \frac{W\big(\widehat{T}_{\mathcal{G}}(\cdot|\mu_1,a), \widehat{T}_{\mathcal{G}}(\cdot|\mu_2,a)\big)}{W(\mu_1,\mu_2)} \leq K_F$$

Proof.

$$\begin{aligned}
W\big(\widehat{T}(\cdot \mid \mu_1,a), \widehat{T}(\cdot \mid \mu_2,a)\big) \quad &:= \quad \inf_j \int_{s_1'} \int_{s_2'} j(s_1',s_2') d(s_1',s_2') ds_1' ds_2' \\
&= \quad \inf_j \int_{s_1} \int_{s_2} \int_{s_1'} \int_{s_2'} \sum_f \mathbb{1}\big(f(s_1)=s_1' \wedge f(s_2)=s_2'\big) j(s_1,s_2,f) d(s_1',s_2') ds_1' ds_2' ds_1 ds_2 \\
&= \quad \inf_j \int_{s_1} \int_{s_2} \sum_f j(s_1,s_2,f) d\big(f(s_1),f(s_2)\big) ds_1 ds_2 \\
&\leq \quad K_F \inf_j \int_{s_1} \int_{s_2} \sum_f g(f|a) j(s_1,s_2) d(s_1,s_2) ds_1 ds_2 \\
&= \quad K_F \sum_f g(f|a) \inf_j \int_{s_1} \int_{s_2} j(s_1,s_2) d(s_1,s_2) ds_1 ds_2 \\
&= \quad K_F \sum_f g(f|a) W(\mu_1,\mu_2) = K_F W(\mu_1,\mu_2)
\end{aligned}$$

Dividing by $W(\mu_1,\mu_2)$ and taking $\sup$ over a, μ_1, and μ_2, we conclude:

$$K^{\mathcal{A}}_{W,W}(\widehat{T}) = \sup_a \sup_{\mu_1,\mu_2} \frac{W\big(\widehat{T}(\cdot \mid \mu_1,a), \widehat{T}(\cdot \mid \mu_2,a)\big)}{W(\mu_1,\mu_2)} \leq K_F \ .$$

We can also prove this using the Kantarovich-Rubinstein duality theorem:

For every μ_1, μ_2, and $a \in \mathcal{A}$ we have:

$$\begin{aligned}
W\big(\widehat{T}_{\mathcal{G}}(\cdot \mid \mu_1,a), \widehat{T}_{\mathcal{G}}(\cdot \mid \mu_2,a)\big) \quad &= \quad \sup_{f:K_{d_{\mathcal{S}},\mathbb{R}}(f)\leq 1} \int_s \big(\widehat{T}_{\mathcal{G}}(s|\mu_1,a) - \widehat{T}_{\mathcal{G}}(s|\mu_2,a)\big) f(s) ds \\
&= \quad \sup_{f:K_{d_{\mathcal{S}},\mathbb{R}}(f)\leq 1} \int_s \int_{s_0} \Big(\widehat{T}(s|s_0,a)\mu_1(s_0) - \widehat{T}(s|s_0,a)\mu_2(s_0)\Big) f(s) ds ds_0 \\
&= \quad \sup_{f:K_{d_{\mathcal{S}},\mathbb{R}}(f)\leq 1} \int_s \int_{s_0} \widehat{T}(s|s_0,a)\Big(\mu_1(s_0) - \mu_2(s_0)\Big) f(s) ds ds_0 \\
&= \quad \sup_{f:K_{d_{\mathcal{S}},\mathbb{R}}(f)\leq 1} \int_s \int_{s_0} \sum_t g(t \mid a) \mathbb{1}\big(t(s_0)=s\big)\Big(\mu_1(s_0) - \mu_2(s_0)\Big) f(s) ds ds_0 \\
&= \quad \sup_{f:K_{d_{\mathcal{S}},\mathbb{R}}(f)\leq 1} \sum_t g(t \mid a) \int_{s_0} \int_s \mathbb{1}\big(t(s_0)=s\big)\big(\mu_1(s_0) - \mu_2(s_0)\big) f(s) ds ds_0 \\
&= \quad \sup_{f:K_{d_{\mathcal{S}},\mathbb{R}}(f)\leq 1} \sum_t g(t \mid a) \int_{s_0} \big(\mu_1(s_0) - \mu_2(s_0)\big) f\big(t(s_0)\big) ds_0
\end{aligned}$$

$$
\begin{aligned}
&\leq \sum_t g(t \mid a) \sup_{f:K_{d_\mathcal{S},\mathbb{R}}(f)\leq 1} \int_{s_0} \big(\mu_1(s_0) - \mu_2(s_0)\big) f\big(t(s_0)\big) ds_0 \\
&\quad \text{composition of } f, t \text{ is Lipschitz with constant upper bounded by } K_F. \\
&= K_F \sum_t g(t \mid a) \sup_{f:K_{d_\mathcal{S},\mathbb{R}}(f)\leq 1} \int_{s_0} \big(\mu_1(s_0) - \mu_2(s_0)\big) \frac{f(t(s_0))}{K_F} ds_0 \\
&\leq K_F \sum_t g(t \mid a) \sup_{h:K_{d_\mathcal{S},\mathbb{R}}(h)\leq 1} \int_{s_0} \big(\mu_1(s_0) - \mu_2(s_0)\big) h(s_0)) ds_0 \\
&= K_F \sum_t g(t \mid a) W(\mu_1, \mu_2) = K_F W(\mu_1, \mu_2)
\end{aligned}
$$

Again we conclude by dividing by $W(\mu_1, \mu_2)$ and taking $\sup$ over a, μ_1, and μ_2. □

Lemma 2. *(Composition Lemma) Define three metric spaces (M_1, d_1), (M_2, d_2), and (M_3, d_3). Define Lipschitz functions $f : M_2 \mapsto M_3$ and $g : M_1 \mapsto M_2$ with constants $K_{d_2,d_3}(f)$ and $K_{d_1,d_2}(g)$. Then, $h : f \circ g : M_1 \mapsto M_3$ is Lipschitz with constant $K_{d_1,d_3}(h) \leq K_{d_2,d_3}(f) K_{d_1,d_2}(g)$.*

Proof.

$$
\begin{aligned}
K_{d_1,d_3}(h) &= \sup_{s1,s2} \frac{d_3\Big(f\big(g(s_1)\big), f\big(g(s_2)\big)\Big)}{d_1(s_1, s_2)} \\
&= \sup_{s_1,s_2} \frac{d_2\big(g(s_1), g(s_2)\big)}{d_1(s_1, s_2)} \frac{d_3\Big(f\big(g(s_1)\big), f\big(g(s_2)\big)\Big)}{d_2\big(g(s_1), g(s_2)\big)} \\
&\leq \sup_{s_1,s_2} \frac{d_2\big(g(s_1), g(s_2)\big)}{d_1(s_1, s_2)} \sup_{s_1,s_2} \frac{d_3\big(f(s_1), f(s_2)\big)}{d_2(s_1, s_2)} \\
&= K_{d_1,d_2}(g) K_{d_2,d_3}(f).
\end{aligned}
$$

□

Lemma 3. *Given a Lipschitz function $f : \mathcal{S} \mapsto \mathbb{R}$ with constant $K_{d_\mathcal{S},d_\mathbb{R}}(f)$:*

$$
K^{\mathcal{A}}_{d_\mathcal{S},d_\mathbb{R}}\Big(\int \widehat{T}(s'|s,a) f(s') ds'\Big) \leq K_{d_\mathcal{S},d_\mathbb{R}}(f) K^{\mathcal{A}}_{d_\mathcal{S},W}\big(\widehat{T}\big) .
$$

Proof.

$$
\begin{aligned}
K^{\mathcal{A}}_{d_\mathcal{S},d_\mathbb{R}}\Big(\int_{s'} \widehat{T}(s'|s,a) f(s') ds'\Big) &= \sup_a \sup_{s_1,s_2} \frac{|\int_{s'} \big(\widehat{T}(s'|s_1,a) - \widehat{T}(s'|s_2,a)\big) f(s') ds'|}{d(s_1,s_2)} \\
&= \sup_a \sup_{s_1,s_2} \frac{|\int_{s'} \big(\widehat{T}(s'|s_1,a) - \widehat{T}(s'|s_2,a)\big) f(s') \frac{K_{d_\mathcal{S},d_\mathbb{R}}(f)}{K_{d_\mathcal{S},d_\mathbb{R}}(f)} ds'|}{d(s_1,s_2)} \\
&= K_{d_\mathcal{S},d_\mathbb{R}}(f) \sup_a \sup_{s_1,s_2} \frac{|\int_{s'} \big(\widehat{T}(s'|s_1,a) - \widehat{T}(s'|s_2,a)\big) \frac{f(s')}{K_{d_\mathcal{S},d_\mathbb{R}}(f)} ds'|}{d(s_1,s_2)} \\
&\leq K_{d_\mathcal{S},d_\mathbb{R}}(f) \sup_a \sup_{s_1,s_2} \frac{|\sup_{g:K_{d_\mathcal{S},d_\mathbb{R}}(g)\leq 1} \int_{s'} \big(\widehat{T}(s'|s_1,a) - \widehat{T}(s'|s_2,a)\big) g(s') ds'|}{d(s_1,s_2)} \\
&= K_{d_\mathcal{S},d_\mathbb{R}}(f) \sup_a \sup_{s_1,s_2} \frac{\sup_{g:K_{d_\mathcal{S},d_\mathbb{R}}(g)\leq 1} \int_{s'} \big(\widehat{T}(s'|s_1,a) - \widehat{T}(s'|s_2,a)\big) g(s') ds'}{d(s_1,s_2)} \\
&= K_{d_\mathcal{S},d_\mathbb{R}}(f) \sup_a \sup_{s_1,s_2} \frac{W\big(\widehat{T}(\cdot|s_1,a), \widehat{T}(\cdot|s_2,a)\big)}{d(s_1,s_2)} \\
&= K_{d_\mathcal{S},d_\mathbb{R}}(f) K^{\mathcal{A}}_{d_\mathcal{S},W}\big(\widehat{T}\big) .
\end{aligned}
$$

$\square$

Lemma 4. *The following operators (Asadi & Littman, 2017) are Lipschitz with constants:*

1. $K_{\|\|_\infty, d_R}(\max(x)) = K_{\|\|_\infty, d_R}\big(mean(x)\big) = K_{\|\|_\infty, d_R}(\epsilon\text{-}greedy(x)) = 1$
2. $K_{\|\|_\infty, d_R}(mm_\beta(x) := \frac{\log \frac{\sum_i e^{\beta x_i}}{n}}{\beta}) = 1$
3. $K_{\|\|_\infty, d_R}(boltz_\beta(x) := \frac{\sum_{i=1}^n x_i e^{\beta x_i}}{\sum_{i=1}^n e^{\beta x_i}}) \leq \sqrt{|A|} + \beta V_{\max}|A|$

Proof. 1 was proven by Littman & Szepesvári (1996), and 2 is proven several times (Fox et al., 2016; Asadi & Littman, 2017; Nachum et al., 2017; Neu et al., 2017). We focus on proving 3. Define

$$\rho(x)_i = \frac{e^{\beta x_i}}{\sum_{i=1}^n e^{\beta x_i}} \ ,$$

and observe that $boltz_\beta(x) = x^\top \rho(x)$. Gao & Pavel (2017) showed that ρ is Lipschitz:

$$\|\rho(x_1) - \rho(x_2)\|_2 \leq \beta \, \|x_1 - x_2\|_2 \tag{1}$$

Using their result, we can further show:

$$\begin{aligned}
&|\rho(x_1)^\top x_1 - \rho(x_2)^\top x_2| \\
&\leq |\rho(x_1)^\top x_1 - \rho(x_1)^\top x_2| + |\rho(x_1)^\top x_2 - \rho(x_2)^\top x_2| \\
&\leq \|\rho(x_1)\|_2 \, \|x_1 - x_2\|_2 \\
&+ \|x_2\|_2 \, \|\rho(x_1) - \rho(x_2)\|_2 \quad \text{(Cauchy-Shwartz)} \\
&\leq \|\rho(x_1)\|_2 \, \|x_1 - x_2\|_2 \\
&+ \|x_2\|_2 \, \beta \, \|x_1 - x_2\|_2 \quad \big(\text{from Eqn 1}\big) \\
&\leq (1 + \beta V_{\max}\sqrt{|A|}) \, \|x_1 - x_2\|_2 \\
&\leq (\sqrt{|A|} + \beta V_{\max}|A|) \, \|x_1 - x_2\|_\infty \ ,
\end{aligned}$$

dividing both sides by $\|x_1 - x_2\|_\infty$ leads to 3. $\square$

Below, we derive the Lipschitz constant for various functions.
ReLu non-linearity We show that ReLu : $\mathbb{R}^n \to \mathbb{R}^n$ has Lipschitz constant 1 for p.

$$\begin{aligned}
K_{\|\cdot\|_p, \|\cdot\|_p}(\text{ReLu}) &= \sup_{x_1, x_2} \frac{\|\text{ReLu}(x_1) - \text{ReLu}(x_2)\|_p}{\|x_1 - x_2\|_p} \\
&= \sup_{x_1, x_2} \frac{(\sum_i |\text{ReLu}(x_1)_i - \text{ReLu}(x_2)_i|^p)^{\frac{1}{p}}}{\|x_1 - x_2\|_p} \\
&\quad \text{(We can show that } |\text{ReLu}(x_1)_i - \text{ReLu}(x_2)_i| \leq |x_{1,i} - x_{2,i}| \text{ and so) :} \\
&\leq \sup_{x_1, x_2} \frac{(\sum_i |x_{1,i} - x_{2,i}|^p)^{\frac{1}{p}}}{\|x_1 - x_2\|_p} \\
&= \sup_{x_1, x_2} \frac{\|x_1 - x_2\|_p}{\|x_1 - x_2\|_p} = 1
\end{aligned}$$

Matrix multiplication Let $W \in \mathbb{R}^{n \times m}$. We derive the Lipschitz continuity for the function $\times W(x) = Wx$.

For $p = \infty$ we have:

$$K_{\|\|_\infty, \|\|_\infty}\big(\times W(x_1)\big)$$

$$
\begin{aligned}
&= \sup_{x_1,x_2} \frac{\left\|\times W(x_1) - \times W(x_2)\right\|_\infty}{\left\|x_1 - x_2\right\|_\infty} = \sup_{x_1,x_2} \frac{\left\|Wx_1 - Wx_2\right\|_\infty}{\left\|x_1 - x_2\right\|_\infty} = \sup_{x_1,x_2} \frac{\left\|W(x_1 - x_2)\right\|_\infty}{\left\|x_1 - x_2\right\|_\infty} \\
&= \sup_{x_1,x_2} \frac{\sup_j \left|W_j(x_1 - x_2)\right|}{\left\|x_1 - x_2\right\|_\infty} \\
&\leq \sup_{x_1,x_2} \frac{\sup_j \left\|W_j\right\| \left\|x_1 - x_2\right\|_\infty}{\left\|x_1 - x_2\right\|_\infty} \quad \text{(Hölder's inequality)} \\
&= \sup_j \left\|W_j\right\|_1 \ ,
\end{aligned}
$$

where W_j refers to jth row of the weight matrix W. Similarly, for $p = 1$ we have:

$$
\begin{aligned}
&K_{\|\|_1,\|\|_1}\big(\times W(x_1)\big) \\
&= \sup_{x_1,x_2} \frac{\left\|\times W(x_1) - \times W(x_2)\right\|_1}{\left\|x_1 - x_2\right\|_1} = \sup_{x_1,x_2} \frac{\left\|Wx_1 - Wx_2\right\|_1}{\left\|x_1 - x_2\right\|_1} = \sup_{x_1,x_2} \frac{\left\|W(x_1 - x_2)\right\|_1}{\left\|x_1 - x_2\right\|_1} \\
&= \sup_{x_1,x_2} \frac{\sum_j \left|W_j(x_1 - x_2)\right|}{\left\|x_1 - x_2\right\|_1} \\
&\leq \sup_{x_1,x_2} \frac{\sum_j \left\|W_j\right\|_\infty \left\|x_1 - x_2\right\|_1}{\left\|x_1 - x_2\right\|_1} = \sum_j \left\|W_j\right\|_\infty \ ,
\end{aligned}
$$

and finally for $p = 2$:

$$
\begin{aligned}
&K_{\|\|_2,\|\|_2}\big(\times W(x_1)\big) \\
&= \sup_{x_1,x_2} \frac{\left\|\times W(x_1) - \times W(x_2)\right\|_2}{\left\|x_1 - x_2\right\|_2} = \sup_{x_1,x_2} \frac{\left\|Wx_1 - Wx_2\right\|_2}{\left\|x_1 - x_2\right\|_2} = \sup_{x_1,x_2} \frac{\left\|W(x_1 - x_2)\right\|_2}{\left\|x_1 - x_2\right\|_2} \\
&= \sup_{x_1,x_2} \frac{\sqrt{\sum_j \left|W_j(x_1 - x_2)\right|^2}}{\left\|x_1 - x_2\right\|_2} \\
&\leq \sup_{x_1,x_2} \frac{\sqrt{\sum_j \left\|W_j\right\|_2^2 \left\|x_1 - x_2\right\|_2^2}}{\left\|x_1 - x_2\right\|_2} = \sqrt{\sum_j \left\|W_j\right\|_2^2} \quad .
\end{aligned}
$$

Vector addition We show that $+b : \mathbb{R}^n \to \mathbb{R}^n$ has Lipschitz constant 1 for $p = 0, 1, \infty$ for all $b \in \mathbb{R}^n$.

$$
\begin{aligned}
K_{\|\cdot\|_p,\|\cdot\|_p}(\text{ReLu}) &= \sup_{x_1,x_2} \frac{\left\| +b(x_1) - +b(x_2)\right\|_p}{\left\|x_1 - x_2\right\|_p} \\
&= \sup_{x_1,x_2} \frac{\left\|(x_1 + b) - (x_2 + b)\right\|_p}{\left\|x_1 - x_2\right\|_p} = \frac{\left\|x_1 - x_2\right\|_p}{\left\|x_1 - x_2\right\|_p} = 1
\end{aligned}
$$

Supervised-learning domain We used the following 5 functions to generate the dataset:

$$
\begin{aligned}
f_0(x) &= \tanh(x) + 3 \\
f_1(x) &= x * x \\
f_2(x) &= \sin(x) - 5 \\
f_3(x) &= \sin(x) - 3 \\
f_4(x) &= \sin(x) * \sin(x)
\end{aligned}
$$

We sampled each function 30 times, where the input was chosen uniformly randomly from $[-2, 2]$ each time.

References

Asadi, K. and Littman, M. L. An alternative softmax operator for reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning*, pp. 243–252, 2017.

Fox, R., Pakman, A., and Tishby, N. G-learning: Taming the noise in reinforcement learning via soft updates. *Uncertainty in Artifical Intelligence*, 2016.

Gao, B. and Pavel, L. On the properties of the softmax function with application in game theory and reinforcement learning. *arXiv preprint arXiv:1704.00805*, 2017.

Littman, M. L. and Szepesvári, C. A generalized reinforcement-learning model: Convergence and applications. In *Proceedings of the 13th International Conference on Machine Learning*, pp. 310–318, 1996.

Nachum, O., Norouzi, M., Xu, K., and Schuurmans, D. Bridging the gap between value and policy based reinforcement learning. *arXiv preprint arXiv:1702.08892*, 2017.

Neu, G., Jonsson, A., and Gómez, V. A unified view of entropy-regularized Markov decision processes. *arXiv preprint arXiv:1705.07798*, 2017.