A. Proof of Claim 2

Proof. For $p = 3$, the desired equation holds, since the matrix $A^T$ becomes just

$$A^T = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix},$$

with characteristic polynomial $(\lambda - 1)\lambda - 1 = \lambda^2 - \lambda - 1$. Let $I$ denote the identity matrix of size $(p - 1) \times (p - 1)$. Assume $p \geq 5$. We consider the matrix:

$$A^T - \lambda I = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

where $A_{11} := \begin{pmatrix} 1 - \lambda & 0 & 0 & 0 & 0 & \ldots & 0 \\ 0 & -\lambda & 0 & 0 & 0 & \ldots & 0 \\ 0 & 0 & -\lambda & 0 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 0 & -\lambda \end{pmatrix}$, $A_{12} := \begin{pmatrix} 1 & 0 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & 0 & \ldots & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 1 \end{pmatrix}$, $A_{21} := \begin{pmatrix} 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & 1 \\ 1 & 1 & 1 & 1 & \ldots & 1 & 1 \end{pmatrix}$, and $A_{22} := \begin{pmatrix} -\lambda & 0 & 0 & \ldots & 0 & 0 \\ 0 & -\lambda & 0 & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 0 & -\lambda \end{pmatrix}$.

Observe that $\lambda = 0$ is not an eigenvalue of the matrix $A^T$. Suppose that $A_{11}, A_{12}, A_{21}, A_{22}$ are the four block submatrices of the matrix above. Using Schur’s complement, we get that $\det(A^T - \lambda I) = \det(A_{22}) \times \det(A_{11} - A_{12}A_{22}^{-1}A_{21})$, where $\det(A_{22}) = (-\lambda)^{\frac{p-1}{2}}$ and

$$\lambda^{\frac{p-1}{2}} \det(A_{11} - A_{12}A_{22}^{-1}A_{21}) = \begin{pmatrix} \lambda - \lambda^2 & 1 & 0 & 0 & \ldots & 0 \\ 0 & -\lambda^2 & 1 & 0 & \ldots & 0 \\ 0 & 0 & -\lambda^2 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 1 & 1 & 1 & \ldots & 1 \end{pmatrix}.$$

We can multiply the first row by $\frac{1}{\lambda(A-1)}$, the second row by $\frac{1}{\lambda^2 + \frac{1}{\lambda^2(A-1)}}$, the third row by $\frac{1}{\lambda^2 + \frac{1}{\lambda^2(A-1)}}$, etc.
We will use induction. Assume now, that we have the result for some $\lambda_1 = \lambda_2 = \ldots = \lambda_p = 0$. Since $\|A^t\|_\infty \geq \text{spec}(A^t) = \rho^t$ (where spec$(A)$ denotes the spectral radius), that is the growth rate of the number of oscillations of compositions of $f$ is at least $\rho_p$. Assume $1 < p$ be an odd number. It suffices to show that $\rho_{p+2} < \rho_p$ (and then use induction). Observe that $\lambda^p - 2\lambda^p - 1 = \lambda^2 - 2\lambda^2 - 1 + \lambda^2$. Therefore

$$0 = q_{p+2}(\rho_{p+2}) = q_{p+2}^2(\rho_{p+2}) + \rho_{p+2}^2 - 1,$$

hence since $\rho_{p+2} > 1$ we conclude that $q_p(\rho_{p+2}) < 0$. Since $\lim_{t \to \infty} q_p(\lambda(t)) = +\infty$, by Bolzano’s theorem it follows that $q_p$ has a root in the interval $(\rho_{p+2}, +\infty)$. Thus $\rho_p > \rho_{p+2}$. One can also see that $\sqrt{2^p - 2\sqrt{2^p - 1}} < 0$ and $2^p - 2 \cdot 2^{p-2} - 1 > 0$, thus from Bolzano’s again, it follows that $\rho_p > \sqrt{2}$ for all $p$.

\[ \sum_{j=1}^{p-3} (-1)^j \lambda^j = -\lambda^p + \lambda^{p-2} + \frac{1}{\lambda + 1} = -\lambda^{p-2} + \frac{1}{\lambda + 1} = -\lambda^{p-2} + \frac{1}{\lambda + 1}, \]

and the claim follows. \[ \square \]

**B. Proof of Corollary 3.5**

**Proof.** We first need to relate the spectral radius with the number of oscillations. We follow the idea from (Chatziafratis et al., 2020) which concludes that $\delta_0 \geq \|A^t\|_\infty \geq \text{spec}(A^t) = \text{spec}(A)^t = \rho^t$ (where spec$(A)$ denotes the spectral radius), that is the growth rate of the number of oscillations of compositions of $f$ is at least $\rho_p$. Assume $1 < p$ be an odd number. It suffices to show that $\rho_{p+2} < \rho_p$ (and then use induction). Observe that $\lambda^p - 2\lambda^p - 1 = \lambda^2 - 2\lambda^2 - 1 + \lambda^2$. Therefore

$$0 = q_{p+2}(\rho_{p+2}) = q_{p+2}^2(\rho_{p+2}) + \rho_{p+2}^2 - 1,$$

hence since $\rho_{p+2} > 1$ we conclude that $q_p(\rho_{p+2}) < 0$. Since $\lim_{t \to \infty} q_p(\lambda(t)) = +\infty$, by Bolzano’s theorem it follows that $q_p$ has a root in the interval $(\rho_{p+2}, +\infty)$. Thus $\rho_p > \rho_{p+2}$. One can also see that $\sqrt{2^p - 2\sqrt{2^p - 1}} < 0$ and $2^p - 2 \cdot 2^{p-2} - 1 > 0$, thus from Bolzano’s again, it follows that $\rho_p > \sqrt{2}$ for all $p$. \[ \square \]

**C. Proof of Lemma 3.6**

**Proof.** It suffices to show that $f$ has period $p$ (the Lipschitz constant is trivially $\rho_p$). We start from $z_0 = 0$ and we get $z_t = f(z_{t-1}) = \rho_p|z_{t-1}| - 1$ for $1 \leq t \leq p$. Observe that $z_1 = -1$, $z_2 = \rho_p - 1 > 0$. Set $q_1(\lambda) = \frac{\lambda - 2\lambda^{p-2} - 1}{\lambda + 1}$. First, we shall show that for $t \in \{3, \ldots, p - 1\}$, we have $z_t \leq 0$ and that $z_t = q_1(\rho_p)$, whereas for $t$ even, we have $z_t = -q_{t-1}(\rho_p) - 1$ in the interval above.

For $t = 3$ we get that $z_3 = \rho_p^2 - \rho_p - 1 = q_3(\rho_p) \leq 0$ because we showed $\rho_p$ is decreasing in $p$ and moreover holds $q_3(\rho_3) = 0$. Since $z_3 \leq 0$ we get that $z_4 = -\rho_p z_3 - 1 = q_3(\rho_p) \rho_p - 1$. Let us show that $z_4 \leq 0$. Observe that $z_4 = -\rho_p^2 + \rho_p^2 + \rho_p - 1 = (\rho_p - 1)(1 - \rho_p^2) < 0$ (since $\rho_p > \sqrt{2}$).

We will use induction. Assume now, that we have the result for some $t$ even, we need to show that $z_{t+1} = q_1(\rho_p)$, $z_{t+2} = -q_{t+1}(\rho_p) - 1$ and moreover $z_{t+1}, z_{t+2} \leq 0$.

By induction, we have that $z_{t+1}, z_{t+2} \leq 0$ and $z_t = -q_{t-1}(\rho_p) \rho_p - 1$, hence $z_{t+1} = -\rho_p(-q_{t-1}(\rho_p) \rho_p - 1) - 1 = \frac{\rho_p^2 - 2\rho_p - 1}{\rho_p + 1} + \rho_p - 1 = q_{t+1}(\rho_p)$. Since $\rho_p$ is decreasing in $p$ and $q_{t+1}(\rho_{t+1}) = 0$, we conclude that $z_{t+1} \leq 0$. Since
We consider three regimes. The first regime corresponds to the functions that appear in Lemma 3.2, where

\[ z_{t+2} = -\rho_p z_{t+1} - 1 = -\rho_p q_{t+1}(\rho_p) - 1. \]

To finish the claim, it suffices to show that \( z_{t+2} \leq 0 \). Observe that

\[
-\rho_p q_{t+1}(\rho_p) - 1 = -\rho_p \left( \rho_p^t - \rho_p^{t-1} - \sum_{j=0}^{t-2} (-\rho_p)^j \right) - 1 \\
= -\rho_p^{t+1} + \rho_p^t - \sum_{j=1}^{t-1} (-\rho_p)^j - 1 \\
= -2\rho_p^{t+1} + 2\rho_p^t + \frac{q_{t+1}(\rho_p)}{\rho_p + 1}.
\]

The term \(-2(\rho_p^{t+1} - \rho_p^t) < 0\) (since \( \rho_p > 1 \)) and moreover \( \frac{q_{t+1}(\rho_p)}{\rho_p + 1} \leq 0 \) because \( \rho_p \) is decreasing in \( p \) and \( t + 1 \leq p - 1 \). Hence \( z_{t+2} \leq 0 \) and the induction is complete.

From the above, we conclude that \( z_p = -\rho_p z_{p-1} - 1 = q_p(\rho_p) = 0 \), thus \( z_0, ..., z_{p-1} \) form a cycle. If we show that \( z_0, ..., z_{p-1} \) are distinct, the proof of the lemma follows.

First observe that \( q_1(\lambda) = \frac{\lambda^{-2} - 2\lambda^{-2} - 1}{\lambda+1} \) is strictly increasing in \( t \) as long as \( \lambda > \sqrt{2} \) (by computing the derivative). Therefore it holds that \( z_3 < z_5 < ... < z_p = 0 \) (for all the odd indices) and also \( z_1 < z_3 \). Furthermore, \( -\lambda q_1(\lambda) - 1 \) is decreasing in \( t \) for \( \lambda > \sqrt{2} \), therefore we conclude \( z_4 > ... > z_{p-1} \) (and also \( z_2 > 0 \).)

We will show that \( z_3 > z_4 \) and finally \( z_{p-1} > -1 = z_1 \) and the lemma will follow. Recall \( z_3 = \rho_p^2 - \rho_p - 1 \) and \( z_4 = -\rho_p^3 + \rho_p^2 + \rho_p - 1 \). Equivalently, we need to show that \( \rho_p^3 - \rho_p - 1 > -\rho_p^3 + \rho_p^2 + \rho_p - 1 \) or \( \rho_p^3 > 2\rho_p - 0 \) which holds because \( \rho_p > \sqrt{2} \). Finally \( z_{p-1} = -\rho_p z_{p-2} - 1 > -1 \) since \( z_{p-2} < z_p = 0 \).

\[ \square \]

D. Sensitivity to Lipschitzness and separation examples based on periods

We consider three regimes. The first regime corresponds to the functions that appear in Lemma 3.2, where \( L = \rho_p \) and \( \rho_p \in [\sqrt{2}, \phi] \), where \( \phi = \frac{1 + \sqrt{5}}{2} \approx 1.618 \) is the golden ratio. The second regime corresponds to the case when \( L > \phi \) and the third regime corresponds to the case when \( L < \sqrt{2} \). We can see in Figure 1 that the function \( f(x) := 2|x| - 1 \) has period 3 and a Lipschitz constant of \( L = 2 \), while in Figure 2, we can see that the function \( f(x) := 1.2|x| - 1 \), does not have any odd period and \( L = 1.2 \).

Figure 1 and Figure 2 correspond to cases where the Lipschitz constant of the function does not match \( \rho_p \).

- When \( \sqrt{2} \leq L \leq \phi \), we see from Figure 3, how small differences in the values of the slope can lead to the existence of different (prime) periods, which consequently lead to different depth-width trade-offs.

- When \( L > \phi \), we can see from Figure 1 that \( L = 2 \) and also the growth rate of oscillations is 2. This means that \( L = \rho \) and that \( L^1 \) separation is achievable. Note that period 3 is present in the tent map, so \( \rho_3 = \phi \) for this case.

- When \( L < \sqrt{2} \), we can see from Figure 2 that the oscillations do not grow exponentially with compositions and that the existing ones are of small magnitude, which means that the \( L^1 \) error can be made arbitrarily small. Observe here that no odd period is present in the function (as this would imply that \( L \geq \rho \geq \sqrt{2} \)).

References

(a) Graph of $f(x)$ intersected with $y = x$, to identify period 1 points.

(b) Graph of $f^3(x)$ intersected with $y = x$, to identify period 3 points.

(c) Graph of $f^5(x)$ intersected with $y = x$, to identify period 5 points.

(d) Graph of $f^7(x)$ intersected with $y = x$, to identify period 7 points.

Figure 1. Here $L = 2$, and this function has period 3. However, the growth rate of oscillations is exactly 2 and since we have equality $L = \rho$ we get $L^1$ separations even though the largest root $\rho_3 = \phi < 2$. 
Better Depth-Width Trade-offs

Figure 2. Here $L = 1.2$ that corresponds to the regime where $L < \sqrt{2}$. It follows that this function cannot have any odd period (because then $L \geq \rho \geq \sqrt{2}$). Observe that the oscillations do not grow exponentially fast and they shrink in area, hence no $L^1$ separation is achievable.
Better Depth-Width Trade-offs

(a) Graph of $f(x)$ is shown. The regime $\sqrt{2} \leq L \leq \phi$ with small slope variations.

(b) Graph of $f^3(x)$. When $L = \phi$, period 3 is present (trade-offs with base $\phi$).

(c) Graph of $f^5(x)$. When $L = 1.513$, period 5 is present (trade-offs with base 1.513).

(d) Graph of $f^7(x)$. When $L = 1.465$, period 7 is present (trade-offs with base 1.465).

Figure 3. Functions parameterized by $\rho_p$ for $L = \rho_p$ and $\rho = 1.618, 1.513, 1.465$ with periods 3, 5 and 7 respectively (see intersection with $y = x$). Slight changes lead to different trade-offs.