# Stability and Chaos in Q-Learning

July 27, 2021

The dynamics we're dealing with are

$$\frac{\dot{x}_i^\mu}{x_i^\mu} = \left( \sum_{\nu \in N_\mu} (Ax^\mu)_i - x^\mu A x^\nu \right) - T(lnx_i^\mu - \langle x^\mu, lnx^\mu \rangle) \tag{1}$$

Now we take the specific case of a two-player game. We can define this in two ways. Either, we choose

$$A = \begin{pmatrix} 0 & r \\ s & 0 \end{pmatrix} \tag{2}$$

or

$$A = \begin{pmatrix} 1 & S \\ T & 0 \end{pmatrix} \tag{3}$$

Both work, but I will start analysing the first case. In this format, if we abuse notation and write $\mathbf{x}^\mu = (x^\mu, 1 - x^\mu)$, the dynamics can be written as

$$\dot{x}^\mu = x_i^\mu (1 - x^\mu) \left( \sum_{\nu \in N_\mu} r - (r+s)x^\nu(t) - Tln\frac{x^\mu}{1-x^\mu} \right) \tag{4}$$

This tells us that a fixed point can only occur where, for each $\mu$, either $x^\mu = 0$, $x^\mu = 1$ or

$$\begin{aligned} &\left( \sum_{\nu \in N_\mu} r - (r+s)x^\nu(t) - Tln\frac{x^\mu}{1-x^\mu} \right) = 0 \\ \implies &\frac{1}{|N_\mu|} \sum_{\nu \in N_\mu} x^\nu(t) = \frac{r}{r+s} - \frac{T}{r+s} ln\frac{x^\mu}{1-x^\mu} \end{aligned} \tag{5}$$

## Interpretation: Allowable Fixed Points at Extreme Temperatures

This immediately tells us something about what fixed points occur at the extremes $T \to 0$ and $T \to \infty$. For (5) to make sense, the right hand side must lie in the range $[0, 1]$. The

reason for this is that the left hand side must be an average over values in $[0, 1]$. For this to make sense, then we must have

$$0 \leq \frac{r}{r+s} - \frac{T}{r+s} ln \frac{x^\mu}{1-x^\mu} \leq 1 \tag{6}$$

which translates to

$$\frac{exp(r/T)}{1+exp(r/T)} \geq x^\mu \geq \frac{exp(-s/T)}{1+exp(-s/T)} \tag{7}$$

We will start by trying to understand the behaviour of Q-Learning close to the fixed point. This can be done by finding the eigenvalues of the Jacobian of the dynamics. Specifically, the Jacobian $J$ is given by

## Linear Stability of Network QL dynamics

$$(J)_{\mu\mu} = \frac{\partial f_\mu}{\partial x^\mu} = (1-2x^\mu) \left[ \sum_{\nu \in N_\mu} r - (r+s)x^\nu \right] - T \left( (1-2x^\mu) \, ln \frac{x^\mu}{1-x^\mu} + 1 \right) \tag{8}$$

$$(J)_{\mu\nu} = \frac{\partial f_\mu}{\partial x^\nu} = \begin{cases} -(r+s)x^\mu(1-x^\mu) & \nu \in N_\mu \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Our job now is to determine the eigenvalues of this Jacobian so that we can assess, on a local scale, the nature of these fixed points. This was done for the case $T = 0$ in the original reference. We now try to understand the effect that T plays on our dynamics. In particular, our goal is to determine whether it is correct to say that chaotic dynamics do not occur in two player games, for any choice of $T$. This was shown to be true for the $T = 0$ case, so we will focus our attention on the $T > 0$ problem. Along the way, we may wish to understand the stability of the fixed points themselves.

### Boundary Fixed Points

For the case $x^\mu = 0$ or $x^\mu = 1$ we have that

$$(J)_{\mu\nu} = -(r+s)x^\mu(1-x^\mu) = 0 \qquad \forall \nu, \tag{10}$$

which means that all of the off-diagonal elements are zero. The eigenvalues are then determined by

$$\lambda_\mu = (1-2x^\mu) \left[ \sum_{\nu \in N_\mu} r - (r+s)x^\nu \right] - T \left( (1-2x^\mu) \, ln \frac{x^\mu}{1-x^\mu} + 1 \right) \tag{11}$$

## Dissipative Dynamics

**Theorem 1.** *Network Q-Learning is a dissipative system for any choice of* $T > 0$.

*Proof.* WRITE PROOF □