# Definitions and Results afferent to Unitary Recurrent Networks

Jean-Philippe Bernardy

April 1, 2022

## 1 Orthogonal matrices

**Definition 1.1** (Orthogonal matrix). *A matrix $Q$ is orthogonal iff it is square and $Q^TQ = I$.*

**Theorem 1.2** (Properties of orthogonal matrices). *A matrix $Q$ is orthogonal iff.*

- $|\det Q| = 1$
- *the rows of $Q$ form an orthogonal basis*
- *the columns of $Q$ form an orthogonal basis*
- *multiplying by $Q$ preserve dot products*
- *multiplying by $Q$ preserve euclidean norms*

**Theorem 1.3** (Orthogonal matrices form a group). *If $Q$ is orthogonal, its inverse is $Q^T$. If $P$ and $Q$ are orthogonal, so is their product $PQ$.*

**Definition 1.4** (Skew-symmetric matrix). *A square matrix $S$ is skew-symmetric iff $S^T = -S$.*

**Theorem 1.5.** *Assume $S$, $u$ and $\lambda$, such that $S^T = -S$ and $Su = \lambda u$. Then $\lambda$ is pure imaginary.*

*Proof.* We have $u^*Su = \lambda u^*u = \lambda\|u\|^2$. Taking the hermitian conjugate of both sides yields $u^*S^*u = \bar{\lambda}\|u\|^2$. Because $S$ is skew-symmetric, we also have $-u^*Su = \bar{\lambda}\|u\|^2$. We then conclude $\lambda = -\bar{\lambda}$, which is satisfied only when $\lambda$ is pure imaginary. $\square$

**Theorem 1.6.** *If $S$ is skew-symmetric then $e^S$ is orthogonal.*

*Proof.* By the spectral theorem, $S$ admits a unitary diagonalisation $S = Q^*\Theta Q$. Let $\Lambda = e^{\Theta}$. By properties of matrix exponential, $e^S = Q^*\Lambda Q$. Because $\Theta$ is diagonal with elements $i\theta_j$, $\Lambda$ is diagonal with elements $\lambda_j = e^{i\theta_j}$. By **??**, every $\theta_j$ is real, and thus every $\lambda_j$ is unimodular. Consequently, $\Lambda$ is unitary. Thus $e^S$ is the product of unitary matrices, and consequently itself unitary. $\square$

## 2 Average effect and distances

**Theorem 2.1.** *For every orthogonal matrix $Q$ of dimension $n$ and a random unit vector $s$, $\mathbb{E}_s[\langle Qs, s\rangle] = \frac{1}{n}\mathsf{trace}(Q)$.*

*Proof.* By the spectral theorem, $Q$ admits a diagonalisation of the form $Q = U^*\Lambda U$, with $U$ unitary. Let $\lambda_i$ be the (diagonal) elements of $\Lambda$ and let $x = Us$. Remark that because $U$ is unitary, $\|x\| = \|s\| = 1$. Thus $\sum_i |x_i|^2 = 1$.[1] Obviously, $\mathbb{E}\left[\sum_i |x_i|^2 = 1\right]$. By assumption, all dimensions of $x$ have the same distribution (applying $Q$ to $s$ does not change this fact, because multiplying by it conserve densities), and thus $\mathbb{E}[|x_i|^2] = \frac{1}{n}$. We can now compute:

$$
\begin{aligned}
\mathbb{E}_s[\langle Qs, s\rangle] &= \mathbb{E}_s[s^T Q s] \\
&= \mathbb{E}_s[s^T U^*\Lambda U s] \\
&= \mathbb{E}_x[x^*\Lambda x] \\
&= \mathbb{E}_x\left[\sum_i |x_i|^2 \lambda_i\right] && \text{by } \Lambda \text{ diagonal} \\
&= \sum_i \lambda_i \mathbb{E}_x[|x_i|^2] && \text{by linearity of expectation} \\
&= \frac{1}{n}\sum_i \lambda_i \\
&= \frac{1}{n}\mathsf{trace}(\Lambda) \\
&= \frac{1}{n}\mathsf{trace}(\Lambda U U^*) \\
&= \frac{1}{n}\mathsf{trace}(U^*\Lambda U)
\end{aligned}
$$

$\square$

**Theorem 2.2.** *For any two orthogonal matrices $P$ and $Q$ of dimension $n$, $\|P - Q\|^2 = 2(n - \langle P, Q\rangle)$.*

[1] Here, even if $Q$ is real, $U$, $\lambda_i$ and thus $x_i$ are complex.

*Proof.*

$$\begin{aligned}\|P-Q\|^2 &= \langle P-Q, P-Q\rangle \\ &= \langle P,P\rangle - \langle P,Q\rangle - \langle P,Q\rangle + \langle Q,Q\rangle \\ &= n - 2\langle P,Q\rangle + n\end{aligned}$$

□

**Theorem 2.3.** *For any two orthogonal matrices $P$ and $Q$ of dimension $n$, and a random unit vector $s$, $\mathbb{E}_s[\langle Ps, Qs\rangle] = \frac{1}{n}\langle P,Q\rangle$.*

*Proof.* By the spectral theorem, $P^TQ$ admits a diagonalisation of the form $P^TQ = U^*\Lambda U$, with $U$ unitary. The proof proceeds as for **??**. □

**Theorem 2.4.** *For every orthogonal matrices $P$ and $Q$ of dimension $n$ and a random unit vector $s$, $\mathbb{E}_s[\|Ps - Qs\|^2] = \frac{1}{n}\|P-Q\|^2$.*

*Proof.* A direct consequence of **??** and **??**. □

# 3 Planes

**Theorem 3.1.** *If $S$ is skew-symmetric, the rank of $S$ gives the maximum number of eigenvalues of $e^S$ different from 1.*

*Proof.* The proof relies on the construction provided in the proof of **??**.

We then note that if $\theta_j = 0$ then $\lambda_j = 1$. Because the rank of $S$ gives the number of non-zero elements of $\Theta$, it is also the maximum number of elements of $\Lambda$ different from 1. (These numbers can differ when $\theta_j = 2\pi$ for some $j$.) □

**Definition 3.2** (Plane similarity)**.** *Assume two planes an $n$-dimensional space, each defined by two orthogonal vectors arranged in a $n$ by $2$ matrix. The similarity between $A$ and $B$ is defined by $\mathsf{sim}(A,B) = \max_\Omega \langle B, A\Omega\rangle$ for $\Omega$ orthogonal.*

The role of taking the minimum for $\Omega$ is to account for equal planes, but which are defined by another basis. Indeed, $A\Omega$ covers all possible bases of the plane defined by $A$ when varying $\Omega$.

**Theorem 3.3.** $\mathsf{sim}(A,B)$ *is the sum of singular values of $B^TA$.*

*Proof.* Let $U\Sigma V^T = B^T A$ be a singular value decomposition of $B^T A$.

$$\begin{aligned}
\max_\Omega \langle B, A\Omega \rangle &= \max_\Omega \mathsf{trace}(B^T A\Omega) \\
&= \max_\Omega \mathsf{trace}(U\Sigma V^T \Omega) \\
&= \max_\Omega \mathsf{trace}(\Sigma V^T \Omega U) \\
&= \mathsf{trace}(\Sigma)
\end{aligned}$$

The last step is justified by $V^T\Omega U$ being orthogonal. Because $\Sigma$ is diagonal, the maximum trace for the product is achieved when $V^T\Omega U = I$. $\square$