Exploring the Time-Domain Solutions to Oja's Equations[1]

February 5, 1997

Oja's equation is a stabilized form of the classical Hebb rule. It describes the learning of the principle components of the input, or in other words, the eigenvectors of the covariance matrix of the input. The Wyatt solution is the full time domain solution of the weight vector, given the covariance matrix of the input and the initial values of the weight vector. In this text we derive the solution, just as Wyatt did, and then use this solution to explore the time dynamics of the neuron in environments modeling the classical visual cortex plasticity experiments.

1 Deriving the Wyatt Equation

This section is just a rewritten form of the first part, and Appendix 1, of Wyatt's original paper. It is here for completeness.

Oja's algorithm can be stated quite simply, with two equations.

$$
\begin{aligned}
y_k &= \mathbf{w_k^T x_k} \text{ (activation rule)} \\
\mathbf{w_{k+1}} &= \mathbf{w_k} + \eta \mathbf{y_k}(\mathbf{x_k} - \mathbf{y_k w_k}) \text{ (learning rule)}
\end{aligned}
$$

where $\mathbf{x_k}, \mathbf{y_k}$ and $\mathbf{w_k}$, are the input vector, output, and weight vector, respectively, at time k.

Taking the limit as $\eta \to 0$, averaging over all input space, assuming that $\langle \mathbf{x} \rangle = \mathbf{0}$, and defining the covariance matrix $\mathbf{C} = \langle \mathbf{x x^T} \rangle$, one arrives at the differential equation for the weight vector:

$$
\dot{\mathbf{w}}(t) = \mathbf{C w(t)} - \mathbf{w(t) w^T(t) C w(t)} \tag{1.1}
$$

We now try to simplify the form of Equation 1.1.

$$
\begin{aligned}
\dot{\mathbf{w}}(t) &= \mathbf{C w(t)} - \mathbf{w(t) w^T(t) C w(t)} \\
&= (\mathbf{C} - \mathbf{w^T C w 1})\mathbf{w} \\
&\equiv \mathbf{A(t) w}
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{A(t)} &\equiv (\mathbf{C} - \mathbf{w^T(t) C w(t) 1}) \equiv (\mathbf{C} - \mathbf{a(t) 1}) \\
a(t) &\equiv \mathbf{w^T(t) C w(t)} \\
\mathbf{1} &\equiv \begin{pmatrix} 1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & 1 \end{pmatrix}
\end{aligned}
$$

Because the quantity $a(t)$ is a scalar, and the covariance matrix $\mathbf{C}$ is time independent, $\mathbf{A(t)}$ has the property $[\mathbf{A(t_1)}, \mathbf{A(t_2)}] = 0$. This lets us immediately write down the solution for $\mathbf{w(t)}$ given the initial condition $\mathbf{w_0}$:

$$
\begin{aligned}
\mathbf{w(t)} &= exp\left[\int_0^t \mathbf{A(t') dt'}\right] \mathbf{w_0} \qquad (1.2) \\
&= e^{\mathbf{Ct}} e^{-\int_0^t a(t') dt'} \mathbf{w_0}
\end{aligned}
$$

[1](J. L. Wyatt and Elfadel, 1995)

Plugging back into the definition of $a(t)$ we get

$$\begin{aligned} a(t) &\equiv \mathbf{w^T(t)Cw(t)} \\ &= e^{-2\int_0^t a(t')dt'} \mathbf{w_0^T Ce^{2Ct} w_0} \end{aligned}$$

Therefore

$$a(t)e^{2\int_0^t a(t')dt'} = \mathbf{w_0^T Ce^{2Ct} w_0}$$

which can be written

$$\frac{d}{dt}\frac{1}{2}e^{2\int_0^t a(t')dt'} = \mathbf{w_0^T \frac{1}{2}\frac{d}{dt}\left[e^{2Ct}\right] w_0}$$

Integrating, and substituting the initial condition, we get

$$e^{2\int_0^t a(t')dt'} = \mathbf{w_0^T e^{2Ct} w_0 + 1 - w_0^T w_0}$$

Substituting this into Equation 1.2 we get the final Wyatt solution:

$$\boxed{\mathbf{w(t)} = \frac{\mathbf{e^{Ct}w_0}}{\left(||\mathbf{e^{Ct}w_0}||^2 + \mathbf{1} - ||\mathbf{w_0}||^2\right)^{\frac{1}{2}}}} \tag{1.3}$$

2 The Classical Rearing Conditions

In this section we hope to use the Wyatt solution to determine the time dynamics of a neuron following Oja's learning rule in an environment modeling the classical visual cortical plasticity experiments. The experiments are called the classical rearing experiments, because the animal is reared in different visual environments. There are 4 such experiments:

- Normal Rearing (NR): animals are brought up normally, with the normal real-world environment as input.
- Monocular Deprivation (MD): starting from a normally reared state, one of the animal's eyes is closed and is thus given uncorrelated input through one eye and normal correlated input through the other eye.
- Binocular Deprivation (BD): starting from a normally reared state, both of the animal's eyes is closed and is thus given uncorrelated input through both eyes.
- Reversed Suture (RS): starting from a monocularly deprived state (final state after MD), the closed eye is now opened, and the previously open eye is closed.

Each of these experiments can be described by the *covariance matrix* of the inputs, and the *initial conditions*. In order to do this, we need to define some notation.

$$\begin{aligned} \text{single eye input vectors:} &\quad \mathbf{x^l} \equiv \text{left eye inputs}, \mathbf{x^r} \equiv \text{right eye inputs} \\ \text{single eye covariance matrix:} &\quad \mathbf{C} \equiv \langle \mathbf{xx^T} \rangle \\ \text{full input vector:} &\quad \mathbf{X} \equiv \begin{pmatrix} \mathbf{x^l} \\ \mathbf{x^r} \end{pmatrix} \end{aligned}$$

full covariance matrix: $\mathcal{C} \equiv \langle \mathbf{X}\mathbf{X}^{\mathbf{T}} \rangle = \left\langle \begin{pmatrix} \mathbf{x^l} \\ \mathbf{x^r} \end{pmatrix} \left((\mathbf{x^l})^{\mathbf{T}} \;\; (\mathbf{x^r})^{\mathbf{T}} \right) \right\rangle$

$= \begin{pmatrix} \mathbf{C^{ll}} & \mathbf{C^{lr}} \\ \mathbf{C^{rl}} & \mathbf{C^{rr}} \end{pmatrix}$

natural scene (single eye) covariance matrix: $\mathbf{C}$

eigenvectors/values of covariance matrix: $\mathbf{C}\mathbf{v_j} = \lambda_{\mathbf{j}}\mathbf{v_j}$ $(\mathbf{j} = \mathbf{1}, \ldots, \mathbf{n})$

$\mathbf{v_i^T}\mathbf{v_j} = \delta_{\mathbf{ij}}$

$\lambda_1 > \lambda_2 > \cdots > \lambda_n$

noise (single eye) covariance matrix: $\boldsymbol{\sigma^2} = \begin{pmatrix} \sigma^2 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \sigma^2 \end{pmatrix}$

Now we can describe the rearing experiments in this new notation.

Experiment	initial weight vector	full covariance matrix
NR	$\mathbf{w_0^{\mathrm{NR}}}$	$\begin{pmatrix} \mathbf{C} & \mathbf{C} \\ \mathbf{C} & \mathbf{C} \end{pmatrix}$
MD	$\mathbf{w_0^{\mathrm{MD}}} = \mathbf{w}^{\mathrm{NR}}(\mathbf{t} \to \infty)$	$\begin{pmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma^2} \end{pmatrix}$
BD	$\mathbf{w_0^{\mathrm{BD}}} = \mathbf{w}^{\mathrm{NR}}(\mathbf{t} \to \infty)$	$\begin{pmatrix} \boldsymbol{\sigma^2} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\sigma^2} \end{pmatrix}$
RS	$\mathbf{w_0^{\mathrm{RS}}} = \mathbf{w}^{\mathrm{MD}}(\mathbf{t} \to \infty)$	$\begin{pmatrix} \boldsymbol{\sigma^2} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix}$

The general procedure will be the following:

- expand the initial weight vector $\mathbf{w_0}$ in terms of the eigenvectors of the covariance matrix ($\mathbf{v_j}$).
- plug into the Wyatt solution (Equation 1.3) to obtain $\mathbf{w}(\mathbf{t})$. We will have to use the full covariance matrix $\mathcal{C}$ instead of the single eye covariance matrix $\mathbf{C}$.
- add approximations to make the equations simpler and to gain some insight into the properties of the dynamics. the approximations may include
  - · $\lambda_1 \gg \lambda_2, ..., \lambda n$: the natural scene covariance matrix is dominated by one eigenvector
  - · $\sigma^2 \ll 1$: noise is small

We present each rearing condition currently.

Normal Rearing (NR)

- expand the initial weight vector $\mathbf{w_0}$ in terms of the eigenvectors of the covariance matrix ($\mathbf{v_j}$).

$$\mathbf{w_0} = \begin{pmatrix} \mathbf{w_0^l} \\ \mathbf{w_0^r} \end{pmatrix} = \sum_{\mathbf{j}} \begin{pmatrix} a_j^l \mathbf{v_j} \\ a_j^r \mathbf{v_j} \end{pmatrix}$$

- plug into the Wyatt solution (Equation 1.3) to obtain $\mathbf{w(t)}$

We need to do this part in steps, calculating $\mathcal{C}\mathbf{w_0}$ first, then calculating each term in Equation 1.3.

$$\begin{aligned}
\mathcal{C}\mathbf{w_0} &= \begin{pmatrix} \mathbf{C} & \mathbf{C} \\ \mathbf{C} & \mathbf{C} \end{pmatrix} \sum_j \begin{pmatrix} a_j^l \mathbf{v_j} \\ a_j^r \mathbf{v_j} \end{pmatrix} = \sum_j \begin{pmatrix} \lambda_j (a_j^l + a_j^r) \mathbf{v_j} \\ \lambda_j (a_j^l + a_j^r) \mathbf{v_j} \end{pmatrix} \\
&= \sum_j \lambda_j (a_j^l + a_j^r) \begin{pmatrix} \mathbf{v_j} \\ \mathbf{v_j} \end{pmatrix} \\
\mathcal{C}^2 \mathbf{w_0} &= \begin{pmatrix} \mathbf{C} & \mathbf{C} \\ \mathbf{C} & \mathbf{C} \end{pmatrix} \sum_j \lambda_j (a_j^l + a_j^r) \begin{pmatrix} \mathbf{v_j} \\ \mathbf{v_j} \end{pmatrix} \\
&= \sum_j 2\lambda_j^2 (a_j^l + a_j^r) \begin{pmatrix} \mathbf{v_j} \\ \mathbf{v_j} \end{pmatrix} \\
&\vdots \\
\mathcal{C}^i \mathbf{w_0} &= \sum_j 2^{i-1} \lambda_j^i (a_j^l + a_j^r) \begin{pmatrix} \mathbf{v_j} \\ \mathbf{v_j} \end{pmatrix} \\
&\vdots
\end{aligned}$$

$$\begin{aligned}
e^{\mathcal{C}t} \mathbf{w_0} &= \left(1 + (\mathcal{C}t) + \frac{(\mathcal{C}t)^2}{2!} + \cdots \right) \mathbf{w_0} \\
&= \sum_j \left\{ (a_j^l + a_j^r) \begin{pmatrix} \mathbf{v_j} \\ \mathbf{v_j} \end{pmatrix} \left(1 + \lambda_j t + \frac{2\lambda_j^2 t^2}{2!} + \cdots + \frac{2^{i-1}\lambda_j^i t^i}{i!} + \cdots \right) - \begin{pmatrix} a_j^r \mathbf{v_j} \\ a_j^l \mathbf{v_j} \end{pmatrix} \right\} \\
&= \sum_j \left\{ \begin{pmatrix} \mathbf{v_j} \\ \mathbf{v_j} \end{pmatrix} (a_j^l + a_j^r)(\frac{1}{2} e^{2\lambda_j t} + \frac{1}{2}) - \begin{pmatrix} a_j^r \mathbf{v_j} \\ a_j^l \mathbf{v_j} \end{pmatrix} \right\} \\
&= \sum_j \frac{1}{2} \begin{pmatrix} \mathbf{v_j} \left[(\mathbf{a_j^l} + \mathbf{a_j^r}) \mathbf{e^{2\lambda_j t}} + (\mathbf{a_j^l} - \mathbf{a_j^r}) \right] \\ \mathbf{v_j} \left[(\mathbf{a_j^l} + \mathbf{a_j^r}) \mathbf{e^{2\lambda t}} + (\mathbf{a_j^r} - \mathbf{a_j^l}) \right] \end{pmatrix}
\end{aligned}$$

$$\begin{aligned}
||\mathbf{w_0}||^2 &= \mathbf{w_0^T} \mathbf{w_0} = \sum_{\mathbf{k}} \sum_{\mathbf{j}} \begin{pmatrix} \mathbf{a_k^l} \mathbf{v_k^T} & \mathbf{a_k^r} \mathbf{v_k^T} \end{pmatrix} \begin{pmatrix} a_j^l \mathbf{v_j} \\ a_j^r \mathbf{v_j} \end{pmatrix} \\
&= \sum_j \left[(a_j^l)^2 + (a_j^r)^2 \right]
\end{aligned}$$

$$\begin{aligned}
a_{j+} &\equiv (a_j^l + a_j^r) \\
a_{j-} &\equiv (a_j^l - a_j^r) \\
\left|\left| e^{\mathcal{C}t} \mathbf{w_0} \right|\right|^2 &= \sum_j \frac{1}{4} \left[a_{j+}^2 e^{4\lambda_j t} + a_{j-}^2 + 2a_{j-}a_{j+}e^{2\lambda_j t} + a_{j+}^2 e^{4\lambda_j t} + a_{j-}^2 - 2a_{j-}a_{j+}e^{2\lambda_j t} \right] \\
&= \frac{1}{2} \sum_j \left[(a_j^l + a_j^r)^2 e^{4\lambda_j t} + (a_j^l - a_j^r)^2 \right]
\end{aligned}$$

which brings us to our final solution

$$\mathbf{w}^{\mathrm{NR}}(\mathbf{t}) = \frac{\sum_{\mathbf{j}} \frac{1}{2} \begin{pmatrix} \mathbf{v_j}\left[(\mathbf{a_j^l}+\mathbf{a_j^r})\mathbf{e^{2\lambda_j t}} + (\mathbf{a_j^l}-\mathbf{a_j^r})\right] \\ \mathbf{v_j}\left[(\mathbf{a_j^l}+\mathbf{a_j^r})\mathbf{e^{2\lambda_j t}} + (\mathbf{a_j^r}-\mathbf{a_j^l})\right] \end{pmatrix}}{\left(\frac{1}{2}\sum_{\mathbf{j}}\left[(\mathbf{a_j^l}+\mathbf{a_j^r})^{\mathbf{2}}\mathbf{e^{4\lambda_j t}} + (\mathbf{a_j^l}-\mathbf{a_j^r})^{\mathbf{2}}\right] + \mathbf{1} - \sum_{\mathbf{j}}\left[(\mathbf{a_j^l})^{\mathbf{2}} + (\mathbf{a_j^r})^{\mathbf{2}}\right]\right)^{\frac{1}{2}}} \tag{2.4}$$

- add approximations to make the equations simpler

In this particular case, we don't need to make any approximations, because the solution we have is tractable and exact. The $t \to \infty$ limiting case, needed for the calculations in the next few sections, is straightforward to calculate. If the largest eigenvalue is non-degenerate, then $\mathbf{w}$ will become

$$\mathbf{w}^{\mathrm{NR}}(\mathbf{t} \to \infty) = \sqrt{\frac{\mathbf{1}}{\mathbf{2}}}\begin{pmatrix} \mathbf{v_1} \\ \mathbf{v_1} \end{pmatrix}$$

Monocular Deprivation (MD)

- expand the initial weight vector $\mathbf{w_0}$ in terms of the eigenvectors of the covariance matrix ($\mathbf{v_j}$).

Since we are starting from the $\mathbf{w}^{\mathrm{NR}}(\mathbf{t} \to \infty)$ state, the initial weight vector for monocular deprivation is already in terms of the eigenvectors of the covariance matrix.

$$\mathbf{w_0} = \sqrt{\frac{\mathbf{1}}{\mathbf{2}}}\begin{pmatrix} \mathbf{v_1} \\ \mathbf{v_1} \end{pmatrix}$$

- plug into the Wyatt solution (Equation 1.3) to obtain $\mathbf{w}(\mathbf{t})$.

$$\begin{aligned}
\mathcal{C}\mathbf{w_0} &= \begin{pmatrix} \mathbf{C} & \mathbf{0} \\ \mathbf{0} & \sigma^{\mathbf{2}} \end{pmatrix}\sqrt{\frac{1}{2}}\begin{pmatrix} \mathbf{v_1} \\ \mathbf{v_1} \end{pmatrix} \\
&= \sqrt{\frac{1}{2}}\begin{pmatrix} \lambda_1\mathbf{v_1} \\ \sigma^2\mathbf{v_1} \end{pmatrix} \\
e^{\mathcal{C}t}\mathbf{w_0} &= \sqrt{\frac{1}{2}}\begin{pmatrix} e^{\lambda_1 t}\mathbf{v_1} \\ e^{\sigma^2 t}\mathbf{v_1} \end{pmatrix} \\
||\mathbf{w_0}||^2 &= 1 \\
\left|\left|e^{\mathcal{C}t}\mathbf{w_0}\right|\right|^2 &= \frac{1}{2}(e^{2\lambda_1 t} + e^{2\sigma^2 t})
\end{aligned}$$

which yields

$$\mathbf{w}^{\mathrm{MD}}(\mathbf{t}) = \frac{\begin{pmatrix} e^{\lambda_1 t}\mathbf{v_1} \\ e^{\sigma^2 t}\mathbf{v_1} \end{pmatrix}}{(\mathbf{e^{2\lambda_1 t}} + \mathbf{e^{2\sigma^2 t}})^{\frac{1}{2}}} \tag{2.5}$$

- add approximations to make the equations simpler

The assumption we are going to make is that $\lambda_1 > \sigma^2$. This is reasonable, because the noise to the closed eye should be smaller than the eigenvalue from the natural scene covariance. With this assumption, the $t \to \infty$ limiting case becomes

$$\mathbf{w}^{\mathrm{MD}}(\mathbf{t} \to \infty) = \begin{pmatrix} \mathbf{v_1} \\ \mathbf{0} \end{pmatrix}$$

Binocular Deprivation (BD)

- expand the initial weight vector $\mathbf{w_0}$ in terms of the eigenvectors of the covariance matrix ($\mathbf{v_j}$).

  As in MD, the initial weight vector is

$$\mathbf{w_0} = \sqrt{\frac{\mathbf{1}}{\mathbf{2}}} \begin{pmatrix} \mathbf{v_1} \\ \mathbf{v_1} \end{pmatrix}$$

- plug into the Wyatt solution (Equation 1.3) to obtain $\mathbf{w(t)}$.

$$\begin{aligned}
\mathcal{C}\mathbf{w_0} &= \begin{pmatrix} \sigma^{\mathbf{2}} & \mathbf{0} \\ \mathbf{0} & \sigma^{\mathbf{2}} \end{pmatrix} \sqrt{\frac{1}{2}} \begin{pmatrix} \mathbf{v_1} \\ \mathbf{v_1} \end{pmatrix} \\
&= \sqrt{\frac{1}{2}} \begin{pmatrix} \sigma^2\mathbf{v_1} \\ \sigma^2\mathbf{v_1} \end{pmatrix} \\
e^{\mathcal{C}t}\mathbf{w_0} &= \sqrt{\frac{1}{2}} \begin{pmatrix} e^{\sigma^2 t}\mathbf{v_1} \\ e^{\sigma^2 t}\mathbf{v_1} \end{pmatrix} \\
||\mathbf{w_0}||^2 &= 1 \\
\left|\left|e^{\mathcal{C}t}\mathbf{w_0}\right|\right|^2 &= e^{2\sigma^2 t}
\end{aligned}$$

which yields

$$\mathbf{w}^{\mathrm{BD}}(\mathbf{t}) = \sqrt{\frac{\mathbf{1}}{\mathbf{2}}} \begin{pmatrix} \mathbf{v_1} \\ \mathbf{v_1} \end{pmatrix} \tag{2.6}$$

Reversed Suture (RS)

- expand the initial weight vector $\mathbf{w_0}$ in terms of the eigenvectors of the covariance matrix ($\mathbf{v_j}$).

We run into an immediate problem if we try to use $\mathbf{w}^{\mathrm{MD}}(\mathbf{t} \to \infty)$ as $\mathbf{w_0}$: the newly opened eye *never* recovers. To alleviate this, we assume that the monocular deprivation experiment did not achieve $t = \infty$, but just some large number T. In that case the initial weight vector for RS is

$$\mathbf{w_0} = \begin{pmatrix} \mathbf{v_1} \\ \epsilon\mathbf{v_1} \end{pmatrix}$$

where $\epsilon \sim e^{(\sigma^2 - \lambda_1)T} \ll 1$

- plug into the Wyatt solution (Equation 1.3) to obtain $\mathbf{w(t)}$.

$$\begin{aligned}
\mathcal{C}\mathbf{w_0} &= \begin{pmatrix} \sigma^{\mathbf{2}} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{v_1} \\ \epsilon\mathbf{v_1} \end{pmatrix} \\
&= \begin{pmatrix} \sigma^2\mathbf{v_1} \\ \lambda_1\epsilon\mathbf{v_1} \end{pmatrix} \\
e^{\mathcal{C}t}\mathbf{w_0} &= \begin{pmatrix} e^{\sigma^2 t}\mathbf{v_1} \\ e^{\lambda_1 t}\epsilon\mathbf{v_1} \end{pmatrix} \\
||\mathbf{w_0}||^2 &= 1 \\
\left|\left|e^{\mathcal{C}t}\mathbf{w_0}\right|\right|^2 &= (e^{2\sigma^2 t} + \epsilon e^{2\lambda_1 t})
\end{aligned}$$

which yields

$$\mathbf{w}^{\mathrm{RS}}(\mathbf{t}) = \frac{\begin{pmatrix} e^{\sigma^2 t}\mathbf{v_1} \\ e^{\lambda_1 t}\epsilon\mathbf{v_1} \end{pmatrix}}{(\mathbf{e}^{\mathbf{2}\sigma^{\mathbf{2}}\mathbf{t}} + \epsilon\mathbf{e}^{\mathbf{2}\lambda_{\mathbf{1}}\mathbf{t}})^{\frac{1}{2}}} \tag{2.7}$$

References

J. L. Wyatt, J. and Elfadel, I. M. (1995). Time-domain solutions of Oja's equations. *Neural Computation*, 7(5):915–922.