To understand the derivation of the Hebbian, anti-Hebbian network written in the paper, I derived it again by myself.

The cost function we would like to optimize is:

$$y_T = \arg\min_{y_T \geq 0} \|X'X - Y'Y\|_F^2 + \lambda\, \mathrm{rank}(Y) \tag{1}$$

$$\begin{aligned}
y_T = &\arg\min_{y_T\geq 0, Card(y_T)\geq Card(y_{T-1})} \left[\sum_{t=1}^{T}\sum_{s=1}^{T}(x_t'x_s - y_t'y_s)^2 + \lambda\, Card(y_T)\right] & (2)\\
= &\arg\min_{y_T\geq 0, Card(y_T)\geq Card(y_{T-1})} \left[2\sum_{t=1}^{T-1}(x_t'x_T - y_t'y_T)^2 + (x_T'x_T - y_T'y_T)^2 + \lambda\, Card(y_T)\right] & (3)\\
= &\arg\min_{y_T\geq 0, Card(y_T)\geq Card(y_{T-1})} \left[2\sum_{t=1}^{T-1}((x_t'x_T)^2 - 2x_t'x_T y_t'y_T + (y_t'y_T)^2) + \right. & (4)\\
&((x_T'x_T)^2 - 2x_T'x_T y_T'y_T + (y_T'y_T)^2) + \\
&\left. + \lambda\, Card(y_T)\right]
\end{aligned}$$

We can remove terms which are not related with y_T, so

$$\begin{aligned}
= &\arg\min_{y_T\geq 0, Card(y_T)\geq Card(y_{T-1})} \left[2\sum_{t=1}^{T-1}(-2x_t'x_T y_t'y_T + (y_t'y_T)^2)\right. & (5)\\
&- 2x_T'x_T y_T'y_T + (y_T'y_T)^2 + \\
&\left. + \lambda\, Card(y_T)\right]\\
= &\arg\min_{y_T\geq 0, Card(y_T)\geq Card(y_{T-1})} \left[\sum_{t=1}^{T-1}(-4x_t'x_T y_t'y_T + 2(y_t'y_T)^2)\right. & (6)\\
&- 2x_T'x_T y_T'y_T + (y_T'y_T)^2 + \\
&\left. + \lambda\, Card(y_T)\right]
\end{aligned}$$

x is n-dim vector and y is m-dim vector.

$$= \underset{y_T \geq 0, Card(y_T) \geq Card(y_{T-1})}{\arg\min} \Bigg[\sum_{t=1}^{T-1} \{(-4 \sum_{j=1}^{n} x_{tj} x_{Tj})(\sum_{i=1}^{m} y_{ti} y_{Ti}) + 2(\sum_{i=1}^{m} y_{ti} y_{Ti})^2\} \tag{7}$$

$$- 2(\sum_{j=1}^{n} x_{Tj} x_{Tj})(\sum_{i=1}^{m} y_{Ti} y_{Ti}) + (\sum_{i=1}^{m} y_{Ti} y_{Ti})^2 +$$

$$+\lambda \, Card(y_T) \Bigg]$$

$$= \underset{y_T \geq 0, Card(y_T) \geq Card(y_{T-1})}{\arg\min} \Bigg[-4 \sum_{i=1}^{m} y_{Ti} \sum_{j=1}^{n} x_{Tj} \sum_{t=1}^{T-1} y_{ti} x_{tj} \tag{8}$$

$$+ 2 \sum_{i=1}^{m} y_{Ti} \sum_{k=1}^{m} y_{Tk} \sum_{t=1}^{T-1} y_{ti} y_{tk}$$

$$- 2 \sum_{i=1}^{m} \parallel x_T \parallel_2^2 \, {y_{Ti}}^2 +$$

$$\sum_{i=1}^{m} {y_{Ti}}^2 \sum_{k=1}^{m} y_{Tk}^2 +$$

$$+\lambda \, Card(y_T) \Bigg]$$

$$= \underset{y_T \geq 0, Card(y_T) \geq Card(y_{T-1})}{\arg\min} \sum_{i=1}^{m} \Bigg[-4 y_{Ti} \sum_{j=1}^{n} x_{Tj} \sum_{t=1}^{T-1} y_{ti} x_{tj} \tag{9}$$

$$+ 2 y_{Ti} \sum_{k=1}^{m} y_{Tk} \sum_{t=1}^{T-1} y_{ti} y_{tk}$$

$$- 2 \parallel x_T \parallel_2^2 \, {y_{Ti}}^2 +$$

$${y_{Ti}}^2 \sum_{k=1}^{m} y_{Tk}^2$$

$$\Bigg] + \lambda \, Card(y_T)$$

We think about optimizing this by applying coordinate descent for each dimension of y_T.

$$
\begin{aligned}
y_{Ti} = & \mathop{\arg\min}_{y_{Ti}\geq 0, Card(y_T)\geq Card(y_{T-1})} \Bigg[-4y_{Ti}\sum_{j=1}^{n} x_{Tj}\sum_{t=1}^{T-1} y_{ti}x_{tj} \\
& + 4y_{Ti}\sum_{k=1,k\neq i}^{m} y_{Tk}\sum_{t=1}^{T-1} y_{ti}y_{tk} + 2y_{Ti}^2\sum_{t=1}^{T-1} y_{ti}^2 \\
& - 2 \parallel x_T \parallel_2^2 {y_{Ti}}^2 + \\
& 2{y_{Ti}}^2 \sum_{k=1,k\neq i}^{m} y_{Tk}^2 + {y_{Ti}}^4\Bigg]
\end{aligned} \tag{10}
$$

$$
\begin{aligned}
= & \mathop{\arg\min}_{y_{Ti}\geq 0, Card(y_T)\geq Card(y_{T-1})} \Bigg[-4y_{Ti}\sum_{j=1}^{n} x_{Tj}\sum_{t=1}^{T-1} y_{ti}x_{tj} \\
& + 4y_{Ti}\sum_{k=1,k\neq i}^{m} y_{Tk}\sum_{t=1}^{T-1} y_{ti}y_{tk} + 2y_{Ti}^2\sum_{t=1}^{T-1} y_{ti}^2 \\
& \left(2\sum_{k=1,k\neq i}^{m} y_{Tk}^2 - 2 \parallel x_T \parallel_2^2\right){y_{Ti}}^2 + {y_{Ti}}^4\Bigg]
\end{aligned} \tag{11}
$$

If the i-th node of output y_T has not been utilized during $t=1$ to $T-1$, i.e. $\sum_{t=1}^{T-1} y_{ti}^2 = 0$, we adjust the output as

$$
y_{Ti} = \mathop{\arg\min}_{y_{Ti}\geq 0}\left[\left(\|x_T\|_2^2 - \sum_{k=1,k\neq i}^{m} y_{Tk}^2\right) - y_{Ti}^2\right]^2 + \lambda\|y_{Ti}\|_0 = \tag{12}
$$
$$
\begin{cases} 0, & \left(\|x_T\|_2^2 - \sum_{k=1,k\neq i}^{m} y_{Tk}^2\right)^2 \leq \lambda \\ \left(\|x_T\|_2^2 - \sum_{k=1,k\neq i}^{m} y_{Tk}^2\right)^{1/2}, & \left(\|x_T\|_2^2 - \sum_{k=1,k\neq i}^{m} y_{Tk}^2\right)^2 > \lambda \end{cases}
$$

Once the i-th output node becomes active, then $\sum_{t=1}^{T-1} y_{ti}^2 > 0$, so can write the equation as

$$
\begin{aligned}
y_{Ti} &= \underset{y_{Ti}\geq 0}{\arg\min} \sum_{t=1}^{T-1} y_{ti}^2 \Bigg[-\frac{4y_{Ti}\sum_{j=1}^{n} x_{Tj}\sum_{t=1}^{T-1} y_{ti}x_{tj}}{\sum_{t=1}^{T-1} y_{ti}^2} \\
&\quad + \frac{4y_{Ti}\sum_{k=1,k\neq i}^{m} y_{Tk}\sum_{t=1}^{T-1} y_{ti}y_{tk}}{\sum_{t=1}^{T-1} y_{ti}^2} + 2y_{Ti}^2 \\
&\quad \frac{2\left(\sum_{k=1,k\neq i}^{m} y_{Tk}^2 - \parallel x_T \parallel_2^2\right){y_{Ti}}^2}{\sum_{t=1}^{T-1} y_{ti}^2} + \frac{{y_{Ti}}^4}{\sum_{t=1}^{T-1} y_{ti}^2} \Bigg] \qquad (13) \\
&= \underset{y_{Ti}\geq 0}{\arg\min} \sum_{t=1}^{T-1} y_{ti}^2 \Bigg[-4y_{Ti}\sum_{j=1}^{n}\left(\frac{\sum_{t=1}^{T-1} y_{ti}x_{tj}}{\sum_{t=1}^{T-1} y_{ti}^2}\right)x_{Tj} \\
&\quad + 4y_{Ti}\sum_{k=1,k\neq i}^{m}\left(\frac{\sum_{t=1}^{T-1} y_{ti}y_{tk}}{\sum_{t=1}^{T-1} y_{ti}^2}\right)y_{Tk} + 2y_{Ti}^2 \\
&\quad \frac{2\left(\sum_{k=1,k\neq i}^{m} y_{Tk}^2 - \parallel x_T \parallel_2^2\right){y_{Ti}}^2}{\sum_{t=1}^{T-1} y_{ti}^2} + \frac{{y_{Ti}}^4}{\sum_{t=1}^{T-1} y_{ti}^2} \Bigg] \qquad (14)
\end{aligned}
$$

At the large T limit, $\sum_{t=1}^{T-1} y_{ti}^2$ in the denominators of the 3rd and 4th term become large. So we can ignore the 3rd and 4th term.

$$
\begin{aligned}
y_{Ti} &\approx \underset{y_{Ti}\geq 0}{\arg\min} \sum_{t=1}^{T-1} y_{ti}^2 \Bigg[-4y_{Ti}\sum_{j=1}^{n}\left(\frac{\sum_{t=1}^{T-1} y_{ti}x_{tj}}{\sum_{t=1}^{T-1} y_{ti}^2}\right)x_{Tj} \qquad (15) \\
&\quad + 4y_{Ti}\sum_{k=1,k\neq i}^{m}\left(\frac{\sum_{t=1}^{T-1} y_{ti}y_{tk}}{\sum_{t=1}^{T-1} y_{ti}^2}\right)y_{Tk} + 2y_{Ti}^2 \Bigg] \\
&\approx \underset{y_{Ti}\geq 0}{\arg\min} \sum_{t=1}^{T-1} y_{ti}^2 \Bigg[-4y_{Ti}\sum_{j=1}^{n} W_{Tij}x_{Tj} \qquad (16) \\
&\quad + 4y_{Ti}\sum_{k=1,k\neq i}^{m} M_{Tik}y_{Tk} + 2y_{Ti}^2 \Bigg]
\end{aligned}
$$

Here we used the substituion of W, M as

$$W_{Tij} = \frac{\sum_{t=1}^{T-1} y_{ti} x_{tj}}{\sum_{t=1}^{T-1} y_{ti}^2} \tag{17}$$

$$M_{Tik \neq i} = \frac{\sum_{t=1}^{T-1} y_{ti} y_{tk}}{\sum_{t=1}^{T-1} y_{ti}^2}; M_{Tii} = 0$$

This optimization can be treated as

$$y_{Ti} \approx \underset{y_{Ti} \geq 0}{\arg\min} (W_{Ti} x_{Tj} - M_{Ti} y_T - y_{Ti})^2 \tag{18}$$

And we obtain,

$$y_{Ti} = \max(W_{Ti} x_T - M_{Ti} y_T, 0) \tag{19}$$