Relation Between Oja's Weighted Subspace Algorithm and Brockett's Algorithm

Terry Sanger, 12/6/93

Oja's Weighted Subspace Algorithm can be written in the averaged form as

$$\dot{Z} \quad = \quad QZ - ZZ^TQZD \tag{1}$$

where $Q = E[xx^T]$ is the autocorrelation matrix of the inputs x, Z is the weight matrix so that the output $y = Z^Tx$, and D is a diagonal matrix with increasing entries.

Brockett's Algorithm in averaged form is

$$\dot{W} \quad = \quad QWN - WNW^TQW \tag{2}$$

where here W is the weight matrix and N is a diagonal matrix with decreasing entries.

Make the change of variables $W = ZD$, $Z = WD^{-1}$ and substitute into equation 1 to obtain

$$\dot{W}D^{-1} = QWD^{-1} - WD^{-2}W^TQW$$

and we see that the this is almost equivalent to Brockett's algorithm if $N = D^{-1}$. Since D has increasing diagonal elements, N will have decreasing diagonal elements, and we have

$$\dot{W}N = QWN - WN^2W^TQW \tag{3}$$

Multiplying on the right by N^{-1} gives

$$\dot{W} = QW - WN^2W^TQWN^{-1} \tag{4}$$

and we can rewrite the last term as $WN(NW^TQWN^{-1})$. Note that $W^TQW = E[yy^T]$ where $y = W^Tx$. Denote this matrix by A and let its entries be $a_{ij} = E[y_iy_j]$. Then we can compare the algorithms

$$\begin{aligned} \dot{W}N^{-1} &= QW - W(NAN^{-1}) && (5)\\ \dot{W} &= QW - WN(NAN^{-1}) && (6)\\ \dot{W} &= QW - W(\mathrm{diag}[A] + 2\mathrm{upper}[A]) && (7)\\ \dot{W} &= QW - W(\mathrm{diag}[A] + \mathrm{upper}[A]) && (8) \end{aligned}$$

where 5 is Brockett's equation, 6 is Oja's Weighted Subspace Algorithm (WSA), 7 is Oja's Stochastic Gradient Ascent (SGA) algorithm, and 8 is Sanger's Generalized Hebbian Algorithm (GHA). "upper[ ]" sets all but the elements strictly above the diagonal to zero, and "diag[ ]" sets all but the diagonal elements to zero. We can look at the elements of the terms in parentheses as follows:

$$\text{Brockett, WSA}: \ = \begin{cases} a_{ij} & i=j \\ (n_i/n_j)a_{ij} & i<j \quad (\text{Note } n_i/n_j > 1) \\ (n_i/n_j)a_{ij} & i>j \quad (\text{Note } n_i/n_j < 1) \end{cases}$$

$$\text{SGA}: \ = \begin{cases} a_{ij} & i=j \\ 2a_{ij} & i<j \\ 0 & i>j \end{cases}$$

$$\text{GHA}: \ = \begin{cases} a_{ij} & i=j \\ a_{ij} & i<j \\ 0 & i>j \end{cases}$$

so Brockett, WSA, and SGA increase the elements of the matrix A above the diagonal, while all four algorithms decrease the elements below the diagonal. The "extra" N term in WSA and N^{-1} term in Brockett's algorithm modify the learning rates to compensate for the factors of n_i/n_j.

We can hypothesize an additional relationship with Oja's Nonlinear PCA algorithm, given in iterative form by

$$\Delta W = xy^T - WyF(y)^T \tag{9}$$

where $F(y)$ applies a nonlinear function $f()$ to each element y_i of the vector y. Suppose that $f()$ is an odd function, and we have a Taylor expansion given by

$$f(y_i) = \alpha_1 y_i + \alpha_3 y_i^3 + \text{h.o.t.}$$

Then ignoring the higher order terms we can rewrite equation 9 as

$$\Delta W = xy^T - Wyy^T(\alpha_1 I + \alpha_3 \text{diag}[y_i^2]) \tag{10}$$

where $\text{diag}[y_i^2]$ is a diagonal matrix with elements y_i^2. If, due to random initial conditions, the expected values $E[y_i^2]$ are distinct, then the averaged form of this equation will be similar to equation 1 with $D = \alpha_1 I + \alpha_3 \text{diag}[y_i^2]$. Although this D is time-varying, only the *order* of its elements is important for convergence.

One problem with this hypothesis is that the element with the smallest value of $E[y_i^2]$ would converge to the eigenvector with largest eigenvalue, so that the output variance would grow and the algorithm appears unstable. We thus require that $\alpha_3 < 0$ and $\alpha_1 > -\alpha_3 \lambda_{\max}$ so that the output variance will decrease with increasing λ_i. Then if $E[y_i^2]$ is smallest at time $t = 0$, w_i will start to converge to the principal eigenvector and $E[y_i^2]$ will remain smallest so that convergence can continue.