A New Training Algorithm for Reservoir Computers

November 5, 2020

Algorithm Description

From initial condition $\mathbf{u}_0$, generate an array $(m \times n_d)$ of samples, U from the system you want to learn. Set the reservoir initial condition with $\mathbf{r}_0 = W_{\text{in}}\mathbf{u}_0$. Drive the reservoir states with the solution array to obtain R, the $(m \times n_r)$ array of reservoir node states.

Solve $||W_{\text{out}}R - U||_2$ for minimizer W_{out}. Tikanov regression with regularization parameter α gives

$$W_{\text{out}} = (R^T R - \alpha I)^{-1} R^T U$$

When n is large the cost of storing R in memory can be prohibitive. This problem is addressed by computing $R^T R$ and $R^T U$ in batches. We write

$$R = \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ \vdots \\ R_k \end{bmatrix}$$

where each Ri is an $(m_i \times n_r)$ array and $\sum_i^k m_i = m$.

Then

$$R^T R = \begin{bmatrix} R_1 & R_2 & R_3 & \cdots & R_k \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \\ R_3 \\ \vdots \\ R_k \end{bmatrix} = \sum_i^k R_i^T R_i$$

If we break up U into

$$U = \begin{bmatrix} U_1 \\ U_2 \\ \vdots \\ U_k \end{bmatrix}$$

where each U_i has dimension $(m_i \times n_d)$ we can compute $R^T U = \sum_i^k R_i^T U_i$. This allows us to compute W_{out} in batches. If the resulting matrices are saved, W_{out} can be updated with additional data.

A challenge with this approach is that the reservoir computer only associates one initial conditions with data. The even if the trained reservoir computer can continue the orbit on which it was trained, it may not learn to predict the orbit of an arbitrary initial condition, even if this new initial condition was close to the original initial condition. This paper presents a solution to this problem.

Continuing with the concept of batch computation of W_{out} we point out that there is no requirement that R_i corresponds with R_j when $j \neq i$. That is, whereas before, we assumed that R was a continuous stream of node states, and W_{out} projects R onto U, what is really true is that W_{out} attempts to send the rows of R to the associated rows in U. Therefore, the matrix U may contain orbits from multiple different initial conditions as long as the corresponding rows of R are the response of the reservoir computer to those orbits and their corresponding initial conditions.

Thus, if U_i contains a discretized solution to an ode, then the first row of R_i should be $W_{\text{in}}\mathbf{u}_0$ where u_0 is the first row of U_i and the remaining rows should be the evolution of node states with input from the entries of U_i. Therefore, each subsection of R must associate with U but the subsections R_i need not relate to each other.

This means that a reservoir computer may be trained with multiple different streams of input. Furthermore, on each individual stream, the reservoir computer can reset it's initial condition so that it learns to associate initial conditions with the appropriate response.

How to break up the data is an important question, because if the batch windows are too small, the reservoir computer will not be trained on long term prediction.

Luckily, since there is no limit on how many rows are in R, we can provide orbit association on multiple time scales if appropriate for the problem at hand. This is done by concatenating different length streams of input data into one large U matrix (Or computing the solution to the Tikhanov Regression problem in batches, with each input data matrix as a separate batch.

Next, focusing on a particular input stream, and assuming it is continuous, we can break it up into overlapping time windows and drive the reservoir computer with each window separately. For each window, we reset the reservoir internal initial condition to correspond with the first input condition of the particular time window. After this we can map the driven internal states back on to the concatenation of time windows. This allows us to train a reservoir computer to replicate an orbit starting from multiple places along the orbit. The number and size of time windows is a hyper parameter that can be tuned.

Algorithm Example

Next we will consider an example of applying this algorithm to learning an ODE. Let's assume we are using a discrete solver for the ODE and therefore can define

a function F which accepts an array of m time values $\mathbf{t} = [t_1, t_2, ..., t_m]$ and an initial condition $\mathbf{u}$ where $\mathbf{u}$ is a vector of length n_d. Passing these values to F produces $U = F(t, \mathbf{u})$ where U is a $(m \times n_d)$ matrix and the i^{th} row of U corresponds to the solution of the ODE at time t_i. Similarly, we can define a function G that corresponds to the untrained reservoir ODE so that for a n_r dimensional vector $\mathbf{r}$, $R = G(t, \mathbf{r}, U)$ where R is an $(m \times n_r)$ matrix and the i^{th} row of R corresponds to the solution to the driven reservoir ode at time t_i.

Algorithm 1: Robust Reservoir Computer Training

Result: Write here the result
$\hat{R} \leftarrow (n_r \times n_r)$ array of zeros
$\hat{U} \leftarrow (n_r \times n_d)$ array of zeros
$T \leftarrow$ Maximum number of time-steps per batch
$\mathbf{v}_1, \mathbf{v}_2, \cdots \mathbf{v}_N \leftarrow$ Initial conditions for the ode
$\tau_1, \tau_2, \cdots \tau_k \leftarrow$ discretized time arrays for each initial condition
for j *in* $1 \cdots N$ **do**
  $\mathbf{t} \leftarrow \tau_j$
  $\mathbf{u} \leftarrow \mathbf{v}_j$
  $U \leftarrow F(\mathbf{t}, \mathbf{u})$
  for i *in* $1 \cdots k$ **do**
    start $\leftarrow T(i-1)+1$
    end $\leftarrow Ti$
    $U_i \leftarrow U$[start:end, :]
    $\mathbf{t}_i \leftarrow \mathbf{t}$[start:end]
    $\mathbf{u}_i \leftarrow U_i[1, :]$
    $\mathbf{r}_0 \leftarrow W_{\text{in}}\mathbf{u}_i$
    $R_i \leftarrow G(\mathbf{t}_i, \mathbf{r}_0, U_i)$
    $\hat{R} \leftarrow \hat{R} + R_i^T R_i$
    $\hat{U} \leftarrow \hat{U} + R_i^T U_i$
  end
end
$W_{\text{out}} \leftarrow (\hat{R} - \alpha I)^{-1}\hat{U}$