Gradient Descent Time Delays Example

Gradient Descent Algorithm

$$x_k = x_{k-1} - \alpha\nabla(f(x_{k-1})) \tag{1}$$

Gradient Descent With Delays

Given delay matrix $D = [d_{ij}] \in \mathbb{N}^{n\times n}$.

$$\begin{bmatrix} x_{k_1} \\ x_{k_2} \\ x_{k_3} \\ \vdots \\ x_{k_n} \end{bmatrix} = \begin{bmatrix} x_{k-1-d_{11_1}} \\ x_{k-1-d_{22_2}} \\ x_{k-1-d_{33_3}} \\ \vdots \\ x_{k-1-d_{nn_n}} \end{bmatrix} - \alpha \begin{bmatrix} \nabla(f(x_{k-1-d_{11_1}}, x_{k-1-d_{12_2}}, \ldots, x_{k-1-d_{1n_n}}))_1 \\ \nabla(f(x_{k-1-d_{21_1}}, x_{k-1-d_{22_2}}, \ldots, x_{k-1-d_{2n_n}}))_2 \\ \nabla(f(x_{k-1-d_{31_1}}, x_{k-1-d_{32_2}}, \ldots, x_{k-1-d_{3n_n}}))_3 \\ \vdots \\ \nabla(f(x_{k-1-d_{n1_1}}, x_{k-1-d_{n2_2}}, \ldots, x_{k-1-d_{nn_n}}))_n \end{bmatrix}$$

Example of Non-Symmetric Delays

Given Delay matrix

$$D = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \tag{2}$$

and function

$$f(x, y) = x^2y + x^2y^2 \tag{3}$$

$$\nabla(f(x, y)) = [2xy + 2xy^2, x^2 + 2x^2y] \tag{4}$$

We will also just define part of a time series so that we can actually see what the delays do. The current time series is the history of states until now. Current time series is

$$T_s = \begin{bmatrix} 3 & 1 \\ 2 & 3 \\ 4 & 5 \\ 1 & 4 \\ 3 & 6 \end{bmatrix}$$

The first update using the delay matrix and time series, (we will select indeces with 1 as the first row and column)

$$\begin{bmatrix} x_{6_1} \\ x_{6_2} \end{bmatrix} = \begin{bmatrix} x_{5-1_1} \\ x_{5-4_2} \end{bmatrix} - \alpha \begin{bmatrix} \nabla(f(x_{5-1_1}, x_{5-2_2}))_1 \\ \nabla(f(x_{5-3_1}, x_{5-4_2}))_2 \end{bmatrix}$$

$$= \begin{bmatrix} x_{4_1} \\ x_{1_2} \end{bmatrix} - \alpha \begin{bmatrix} \nabla(f(x_{4_1}, x_{3_2}))_1 \\ \nabla(f(x_{2_1}, x_{1_2}))_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \alpha \begin{bmatrix} \nabla(f(1, 5))_1 \\ \nabla(f(2, 1))_2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \alpha \begin{bmatrix} [60, 11]_1 \\ [8, 12]_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \alpha \begin{bmatrix} 60 \\ 12 \end{bmatrix}$$

example of symmetric delays

Everything is the same as the previous example except

$$D = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix} \tag{5}$$

The first update from the current time series

$$\begin{bmatrix} x_{6_1} \\ x_{6_2} \end{bmatrix} = \begin{bmatrix} x_{5-1_1} \\ x_{5-2_2} \end{bmatrix} - \alpha \begin{bmatrix} \nabla(f(x_{5-1_1}, x_{5-2_2}))_1 \\ \nabla(f(x_{5-1_1}, x_{5-2_2}))_2 \end{bmatrix}$$

$$= \begin{bmatrix} x_{4_1} \\ x_{3_2} \end{bmatrix} - \alpha \nabla(f(x_{4_1}, x_{3_2})) = \begin{bmatrix} 1 \\ 5 \end{bmatrix} - \alpha \nabla(f(1,5))$$

$$= \begin{bmatrix} 1 \\ 5 \end{bmatrix} - \alpha \begin{bmatrix} 60 \\ 11 \end{bmatrix}$$