Gradient Descent Time Delays Example

Gradient Descent Algorithm

$$x = x - \alpha\nabla(f(x)) \tag{1}$$

Gradient Descent With Delays

Given delay matrix $D = [d_{ij}] \in \mathbb{N}^{n\times n}$. Defining $x_{i,d}$ as the state where i is the index of interaction and d is the amount of delay for that interaction.

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_{1,d_{11}} \\ x_{2,d_{22}} \\ x_{3,d_{33}} \\ \vdots \\ x_{n,d_{nn}} \end{bmatrix} - \alpha \begin{bmatrix} \nabla(f(x_{1,d_{11}}, x_{2,d_{12}}, \dots, x_{n,d_{1n}}))_1 \\ \nabla(f(x_{1,d_{21}}, x_{2,d_{22}}, \dots, x_{n,d_{2n}}))_2 \\ \nabla(f(x_{1,d_{31}}, x_{2,d_{32}}, \dots, x_{n,d_{3n}}))_3 \\ \vdots \\ \nabla(f(x_{1,d_{n1}}, x_{2,d_{n2}}, \dots, x_{n,d_{nn}}))_n \end{bmatrix}$$

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. Current time series, which is all the past states until now, 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)

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_{1,1} \\ x_{2,4} \end{bmatrix} - \alpha \begin{bmatrix} \nabla(f(x_{1,1}, x_{2,2}))_1 \\ \nabla(f(x_{1,3}, x_{2,4}))_2 \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \alpha \begin{bmatrix} \nabla(f(1,5))_1 \\ \nabla(f(2,1))_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \alpha \begin{bmatrix} [60, 11]_1 \\ [8, 12]_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - \alpha \begin{bmatrix} 60 \\ 12 \end{bmatrix}$$

Example of Symmetric Delays

Everything is the same as the previous example except

$$D = \begin{bmatrix} 3 & 4 \\ 3 & 4 \end{bmatrix} \tag{5}$$

The first update from the current time series

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_{1,3} \\ x_{2,4} \end{bmatrix} - \alpha \begin{bmatrix} \nabla(f(x_{1,3}, x_{2,4}))_1 \\ \nabla(f(x_{1,3}, x_{2,4}))_2 \end{bmatrix}$$

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

$$= \begin{bmatrix} 2 \\ 1 \end{bmatrix} - \alpha \begin{bmatrix} 8 \\ 12 \end{bmatrix}$$