Gradient Descent Algorithm

Relevant Formulas

Notation

- The function we want to minimize is denoted by f.
- The number of independent variables of f is denoted by n.
- The vector whose components are the independent variables of f is written as

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

- The algorithm produces a sequence of such vectors. This sequence is denoted by $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$

Iteration Formulas

- Algorithm with a **constant step size**:
  - In this case, we have a single value γ for the step size.
  - We can generate the sequence $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$ by using the equation

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} - \gamma \nabla f \left[\mathbf{x}^{(i)}\right], \quad i = 0, 1, 2, \dots,$$

    where $\mathbf{x}^{(0)}$ must be passed as an input.
- Algorithm with an **adaptive step size**:
  - In this case, we have a sequence of values for the step size, i.e., $\gamma_0, \gamma_1, \gamma_2, \dots$
  - We can generate the sequence $\mathbf{x}^{(0)}, \mathbf{x}^{(1)}, \mathbf{x}^{(2)}, \dots$ by using the equation

$$\mathbf{x}^{(i+1)} = \mathbf{x}^{(i)} - \gamma_i \nabla f \left[\mathbf{x}^{(i)}\right], \quad i = 0, 1, 2, \dots,$$

    where $\mathbf{x}^{(0)}$ and γ_0 must be passed as inputs.
  - The step size γ_i can be written in terms of $\mathbf{x}^{(i)}$ and $\mathbf{x}^{(i-1)}$ as

$$\gamma_i = \frac{\left|\left[\mathbf{x}^{(i)} - \mathbf{x}^{(i-1)}\right]^T \left\{\nabla f \left[\mathbf{x}^{(i)}\right] - \nabla f \left[\mathbf{x}^{(i-1)}\right]\right\}\right|}{\left|\nabla f \left[\mathbf{x}^{(i)}\right] - \nabla f \left[\mathbf{x}^{(i-1)}\right]\right|^2}, \quad i = 1, 2, \dots$$

Example

Using Python, I implemented both versions of the gradient descent algorithm.
To test my code, I chose to minimize the function

$$f(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{n} (x_i - i)^2.$$

It is straightforward to show that the gradient of this function is

$$\nabla f(\mathbf{x}) = \begin{bmatrix} x_1 - 1 \\ x_2 - 2 \\ \vdots \\ x_n - n \end{bmatrix}.$$

This equation allows us to conclude that the minimum of $f(\mathbf{x})$ is at

$$\mathbf{x} = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ n \end{bmatrix}.$$