Algorithms for Numerical Optimization

The problem is to find \( x_m \) that satisfies

\[
f(x_m) = \min_{x \in \mathbb{R}^n} f(x)
\]

Outline of the Steepest Descent Algorithm (SDM)

To use the SDM to minimize the function \( f(x) \) do the following:

1) pick \( x_1 \) and set \( d_1 = -\nabla f(x_1) \)

2) for \( k = 1, 2, 3, \ldots \)
   a) find \( s_k \) that minimizes \( g(s) = f(x_k + sd_k) \)
   b) \( x_{k+1} = x_k + s_k d_k \)
   c) \( d_{k+1} = -\nabla f(x_{k+1}) \)

Outline of the Conjugate Gradient Algorithm (CGM)

To use the CGM to minimize the function \( f(x) \) do the following:

1) pick \( x_1 \) and set \( g_1 = \nabla f(x_1) \) and \( d_1 = -g_1 \)

2) for \( k = 1, 2, 3, \ldots \)
   a) find \( s_k \) that minimizes \( g(s) = f(x_k + sd_k) \)
   b) \( x_{k+1} = x_k + s_k d_k \)
   c) \( g_{k+1} = \nabla f(x_{k+1}) \)
   d) \( d_{k+1} = -g_{k+1} + \beta_k d_k \)
      where \( \beta_k = \frac{g_{k+1} \cdot g_{k+1}}{g_k \cdot g_k} \)
Outline of Newton's Method

To use Newton's method to minimize the function $f(x)$ do the following:

1) pick $x_1$

2) for $k = 1, 2, 3, ...$
   \[ x_{k+1} = x_k - H_k^{-1}\nabla f(x_k) \]
   where
   \[
   H = \begin{bmatrix}
   \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\
   \frac{\partial^2 f}{\partial x_1 \partial x_2} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\
   \vdots & \vdots & \ddots & \vdots \\
   \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2}
   \end{bmatrix}
   \]

Observations on above methods:

1. Both the SDM and CGM are descent methods. This means the value of $f(x)$ decreases at each step but it does not necessarily mean that $x_{k+1}$ is always closer to the solution than $x_k$ (this is true even though the methods converge to the solution).
2. Newton is not a descent method so $f(x)$ does not necessarily decrease at each step. Similarly, it is not necessarily true that $x_{k+1}$ is closer to the solution than $x_k$ is.
3. The convergence of the SDM and CGM methods are:
   a) better than Newton in the sense that they are more likely to converge and the points they converge to are local min points (for convergence Newton must start close to the solution and will converge to local max as well as local min points)
   b) worse than Newton in the sense that they are slower