Method of (Steepest) Gradient Descent

The theory (very short excerpts from lectures)

**Theorem:** Suppose $A$ is a symmetric and positive definite (spd) matrix, $b$ is a vector and $J(x)$ is the quadratic functional $J(x) = \frac{1}{2} x^T A x - x^T b$. Then $A\bar{x} = b \iff J(\bar{x}) < J(x) \quad \forall \; x \neq \bar{x}$.

This Theorem says that the solution of a linear system $Ax = b$ with spd matrix can be found by minimizing the quadratic functional $J(x)$. To achieve this, gradient methods can be used. The most illustrative method of this class is the Method of Gradient Descent, sometimes also called Method of Steepest Descent.

Method of Steepest Descent:

**The main idea:** Start at some point $x_0$, find the direction of the steepest descent of the value of $J(x)$ and move in that direction as long as the value of $J(x)$ descends. At this point, find the new direction of the steepest descent and repeat the whole process.

Note: a direction of the steepest descent of a function at a given point is the direction opposite to its gradient at that point. The gradient is perpendicular to a contour line passing through the given point. See illustration in Figure 1.

Remark: The direction opposite to the gradient of $J(x)$ is equal to the residual $r = b - Ax$ of the system $Ax = b$.

**The algorithm:**
Choose $x^{(0)}$. For $k = 0, 1, 2, \ldots$ compute

1. $r^{(k)} = b - Ax^{(k)}$
2. $\alpha_k = (r^{(k)})^T r^{(k)}/(r^{(k)})^T A r^{(k)}$
3. $x^{(k+1)} = x^{(k)} + \alpha_k r^{(k)}$

until $||r^{(k)}|| < \epsilon$ for a small $\epsilon$ of your choice.

**Example**
Consider a linear system $Ax = b$, where

$$A = \begin{bmatrix} 3 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix}, \quad b = \begin{bmatrix} -1 \\ 7 \\ -7 \end{bmatrix}$$

a) Can the method of steepest descent be used for solving this system?

b) If yes, compute first three iterations by this method, starting from $x^{(0)} = (0, 0, 0)^T$. 

\[b\]
Solution:

a) Let us verify the sufficient condition for using the method. We have to check, if matrix $A$ is spd: $A$ is symmetric, so let us check positive definitness:

$$\det(3) = 3 > 0 \quad , \quad \det \begin{bmatrix} 3 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix} = 20 > 0$$

All leading principal minors are positive and so the matrix $A$ is positive definite.

Conclusion: the method of steepest descent can be used to solve this system.

b) $k = 0$:

1. 

$$r^{(0)} = b - A x^{(0)} = \begin{bmatrix} -1 \\ 7 \\ -7 \end{bmatrix} - \begin{bmatrix} 3 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 7 \\ -7 \end{bmatrix}$$

2. 

$$\begin{bmatrix} -1 \\ 7 \\ -7 \end{bmatrix} \begin{bmatrix} -1 \\ 7 \\ -7 \end{bmatrix} = 1 + 49 + 49 = 99$$

$$\begin{bmatrix} -1 \\ 7 \\ -7 \end{bmatrix} \begin{bmatrix} 3 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} -1 \\ 7 \\ -7 \end{bmatrix} = \begin{bmatrix} -17 \\ 29 \\ -29 \end{bmatrix}$$

$$\alpha_0 = \frac{(r^{(0)})^T r^{(0)}}{(r^{(0)})^T A r^{(0)}} = \frac{99}{423} = 0.2340$$

3. 

$$x^{(1)} = x^{(0)} + \alpha_0 r^{(0)} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} + 0.2340 \begin{bmatrix} -1 \\ 7 \\ -7 \end{bmatrix} = \begin{bmatrix} -0.2340 \\ 1.6383 \\ -1.6383 \end{bmatrix}$$

$k = 1$:

1. 

$$r^{(1)} = b - A x^{(1)} = \begin{bmatrix} -1 \\ 7 \\ -7 \end{bmatrix} - \begin{bmatrix} 3 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} -0.2340 \\ 1.6383 \\ -1.6383 \end{bmatrix} = \begin{bmatrix} 2.9787 \\ 0.2128 \\ -0.2128 \end{bmatrix}$$

2. 

$$\begin{bmatrix} 2.9787 & 0.2128 & -0.2128 \end{bmatrix} \begin{bmatrix} 2.9787 \\ 0.2128 \\ -0.2128 \end{bmatrix} = 8.9633$$

$$\begin{bmatrix} 2.9787 & 0.2128 & -0.2128 \end{bmatrix} \begin{bmatrix} 3 & -1 & 1 \\ -1 & 3 & -1 \\ 1 & -1 & 3 \end{bmatrix} \begin{bmatrix} 2.9787 \\ 0.2128 \\ -0.2128 \end{bmatrix} = \begin{bmatrix} 2.9787 \\ 0.2128 \\ -0.2128 \end{bmatrix}$$

2
Method of (Steepest) Gradient Descent

\[
\begin{bmatrix}
2.9787 & 0.2128 & -0.2128 \\
8.5106 & -2.1277 & 2.1277 \\
\end{bmatrix} = 24.4455
\]

\[\alpha_1 = \frac{(r^{(1)})^T r^{(1)}}{(r^{(1)})^T A r^{(1)}} = \frac{8.9633}{24.4455} = 0.3667\]

3.

\[x^{(2)} = x^{(1)} + \alpha_1 r^{(1)} = \begin{bmatrix}
-0.2340 \\
1.6383 \\
-1.6383 \\
\end{bmatrix} + 0.3667 \begin{bmatrix}
2.9787 \\
0.2128 \\
-0.2128 \\
\end{bmatrix} = \begin{bmatrix}
0.8582 \\
1.7163 \\
-1.7163 \\
\end{bmatrix}\]

\[k = 2:\]

1.

\[r^{(2)} = b - Ax^{(2)} = \begin{bmatrix}
-1 \\
7 \\
-7 \\
\end{bmatrix} - \begin{bmatrix}
3 & -1 & 1 \\
-1 & 3 & 1 \\
1 & -1 & 3 \\
\end{bmatrix} \begin{bmatrix}
0.8582 \\
1.7163 \\
-1.7163 \\
\end{bmatrix} = \begin{bmatrix}
-0.1418 \\
0.9929 \\
-0.9929 \\
\end{bmatrix}\]

2.

\[(r^{(2)})^T r^{(2)} = \begin{bmatrix}
-0.1418 & 0.9929 & -0.9929 \\
-0.1418 & 0.9929 & -0.9929 \\
-0.1418 & 0.9929 & -0.9929 \\
\end{bmatrix} = 1.9919\]

\[(r^{(2)})^T A r^{(2)} = \begin{bmatrix}
3 & -1 & 1 \\
-1 & 3 & 1 \\
1 & -1 & 3 \\
\end{bmatrix} \begin{bmatrix}
-0.1418 \\
0.9929 \\
-0.9929 \\
\end{bmatrix} = \begin{bmatrix}
2.4113 \\
4.1135 \\
-4.1135 \\
\end{bmatrix} = 8.5106
\]

\[\alpha_2 = \frac{(r^{(2)})^T r^{(2)}}{(r^{(2)})^T A r^{(2)}} = \frac{1.9919}{8.5106} = 0.2340\]

3.

\[x^{(3)} = x^{(2)} + \alpha_2 r^{(2)} = \begin{bmatrix}
0.8582 \\
1.7163 \\
-1.7163 \\
\end{bmatrix} + 0.2340 \begin{bmatrix}
-0.1418 \\
0.9929 \\
-0.9929 \\
\end{bmatrix} = \begin{bmatrix}
0.8250 \\
1.9487 \\
-1.9487 \\
\end{bmatrix}\]

\[r^{(3)} = b - Ax^{(3)} = \begin{bmatrix}
-1 \\
7 \\
-7 \\
\end{bmatrix} - \begin{bmatrix}
3 & -1 & 1 \\
-1 & 3 & 1 \\
1 & -1 & 3 \\
\end{bmatrix} \begin{bmatrix}
0.8250 \\
1.9487 \\
-1.9487 \\
\end{bmatrix} = \begin{bmatrix}
0.4225 \\
0.0302 \\
-0.0302 \\
\end{bmatrix}\]

The convergence is quite slow - the exact solution is \(x = (1, 2, -2)^T\).
Figure 1: The ellipses represent contour lines of a quadratic functional. The polygonal line starting at the big bullet (on the outermost ellipse) is a path to the lower values of the functional computed by method of steepest descent.