Numerical Representations Summary

For the M mapping of the genomic sequence S of length N, the following statements apply:

Voss:
$$B(x)=\begin{cases}(1,0,0,0) & \text{if } S(x)=A \\ (0,1,0,0) & \text{if } S(x)=T \\ (0,0,1,0) & \text{if } S(x)=C \\ (0,0,0,1) & \text{if } S(x)=G\end{cases} \quad, x\in[0,\ N-1],\ \ M=B^T$$

Tetrahedron
with $B_0^T=A_n,\ B_1^T=T_n,\ B_2^T=C_n,\ B_3^T=G_n,$ *where*
$$M_r=\frac{\sqrt{2}}{3}(2T_n-C_n-G_n)$$
$$M_g=\frac{\sqrt{6}}{3}(C_n-G_n)$$
$$M_b=\frac{1}{3}(3A_n-T_n-C_n-G_n)$$
$$M(x)=(M_r,M_g,M_b)(x)$$

Chaos Game Representation (CGR)
$$P_0=(0.5,0.5)$$
$$P_x=0.5(P_{x-1}+M(S(x))),\text{where}$$
$$M(S(x))=\begin{cases}(0,0) & \text{if } S(x)=A \\ (1,0) & \text{if } S(x)=T \\ (0,1) & \text{if } S(x)=C \\ (1,1) & \text{if } S(x)=G\end{cases} \quad, x\in[0,\ N-1]$$

Integer CGR
$$P_1=M(S(x))$$
$$P_n=P_{n-1}+2^{(n-1)}M(S(n-1)),\text{ where n = x + 1, and}$$
$$M(S(x))=\begin{cases}(1,1) & \text{if } S(x)=A \\ (-1,1) & \text{if } S(x)=T \\ (-1,-1) & \text{if } S(x)=C \\ (1,-1) & \text{if } S(x)=G\end{cases} \quad, x\in[1,\ N]$$

DNA Walk
$$P_x=P_{x-1}+M(S(x)),where$$
$$M(S(x))=\begin{cases}1 & \text{if } S(x)=A\ \vee\ S(x)=G \\ -1 & \text{if } S(x)=C\ \vee\ S(x)=G\end{cases} \quad, x\in[0,\ N-1]$$

Paired Numeric Mapping:

$$M(x)=\begin{cases} 1 & \text{if } S(x)=A \ \lor \ S(x)=T \\ -1 & \text{if } S(x)=C \ \lor \ S(x)=G \end{cases} \quad ,x\in[0, \ N-1]$$

Complex:

$$M(x)=\begin{cases} 1+i & \text{if } S(x)=A \\ 1-i & \text{if } S(x)=T \\ -1+i & \text{if } S(x)=C \\ -1-i & \text{if } S(x)=G \end{cases} \quad ,x\in[0, \ N-1]$$

Eiip:

$$M(x)=\begin{cases} 0.1260 & \text{if } S(x)=A \\ 0.1335 & \text{if } S(x)=T \\ 0.1340 & \text{if } S(x)=C \\ 0.0806 & \text{if } S(x)=G \end{cases} \quad ,x\in[0, \ N-1]$$

Atomic:

$$M(x)=\begin{cases} 70 & \text{if } S(x)=A \\ 66 & \text{if } S(x)=T \\ 58 & \text{if } S(x)=C \\ 78 & \text{if } S(x)=G \end{cases} \quad ,x\in[0, \ N-1]$$

2-bit binary:

$$M(x)=\begin{cases} 00 & \text{if } S(x)=A \\ 01 & \text{if } S(x)=T \\ 11 & \text{if } S(x)=C \\ 10 & \text{if } S(x)=G \end{cases} \quad ,x\in[0, \ N-1]$$

4-bit binary:

$$M(x)=\begin{cases} 0100 & \text{if } S(x)=A \\ 1000 & \text{if } S(x)=T \\ 0001 & \text{if } S(x)=C \\ 0010 & \text{if } S(x)=G \end{cases} \quad ,x\in[0, \ N-1]$$

Integer:

$$M(x)=\begin{cases} 2 & \text{if } S(x)=A \\ 0 & \text{if } S(x)=T \\ 1 & \text{if } S(x)=C \\ 3 & \text{if } S(x)=G \end{cases} \quad ,x\in[0, \ N-1]$$

Molecular Mass:

$$M(x)=\begin{cases} 134 & \text{if } S(x)=A \\ 125 & \text{if } S(x)=T \\ 110 & \text{if } S(x)=C \\ 150 & \text{if } S(x)=G \end{cases} \quad ,x\in[0, \ N-1]$$

Real:

$$M(x)=\begin{cases}-1.5 & \text{if } S(x)=A \\ 1.5 & \text{if } S(x)=T \\ 0.5 & \text{if } S(x)=C \\ -0.5 & \text{if } S(x)=G \end{cases} \quad ,x\in[0,\ N-1]$$

For the representations Z Curve (Original publication) and Liao 2D graphical representation in the sequence S of size N, whether A_n, T_n, C_n, G_n, the accumulated occurrences of nitrogenous bases A, T, C, G, respectively, up to position n .

The Z curve is defined as a set of nodes P_0, P_1,…, P_n,…, P_N, whose x_n, y_n, z_n coordinates are described below:

Z Curve

$$\begin{cases} A_0=T_0=C_0=G_n=0 \\ x_n=(A_n+G_n)-(C_n+T_n) \\ y_n=(A_n+C_n)-(G_n+T_n) \\ z_n=(A_n+T_n)-(G_n+C_n) \end{cases} \quad ,n\in[0,N]$$

The 2D graphical representation of Liao assumed two constants, defined as n and m, which in Seqreppy are assigned to the values 1/2 and ¾, respectively. Thus, the coordinates of the method's constituent points are defined as:

Liao

$$\begin{cases} x_0=y_0=0 \\ x_i=A_n m + G_n\sqrt{n} + C_n\sqrt{n} + T_n m \\ y_i=-A_n\sqrt{n} - G_n m + C_n m + T_n\sqrt{n} \end{cases} \quad ,i\in[1,N]$$

In Seqreppy they are simplified as follows:

$$\begin{cases} x_0=y_0=0 \\ x_i=m(A_n+T_n) + \sqrt{n}(G_n+C_n) \\ y_i=\sqrt{n}(T_n-A_n) + m(C_n-G_n) \end{cases} \quad ,i\in[1,N]$$