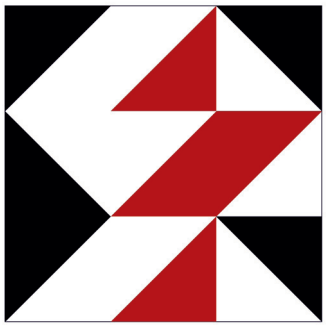


MATHMOD
VIENNA 2012

PREPRINTS
MATHMOD 2012 VIENNA

Abstract Volume

I.Troch, F.Breitenecker, Eds.

7th Vienna Conference on
Mathematical Modelling

February 14-17, 2012
Vienna University of Technology

FRACTIONAL-ORDER FOURIER ANALYSIS OF HUMAN DNA

J. A. Tenreiro Machado
Institute of Engineering of Porto, Dept. of Electrical Engineering, Portugal

Corresponding author: J. A. Tenreiro Machado, Institute of Engineering of Porto, Dept. of Electrical Engineering
Rua Dr. Antonio Bernardino de Almeida, 431, 4200-072 Porto, Portugal, jtm@isep.ipp.pt

Introduction. Phylogenetics is the study of the evolutionary relations between groups of organisms. With the advent of genome sequencing and genome databases, considerable information is available for computational processing, allowing worldwide research on decoding and understanding the informational structure present on DNA sequences.

In this paper is adopted the Fourier analysis and several approximations based on power law concepts. These ideas are usual in Fractional Calculus (FC) which is a paradigm that embodies the standard differential calculus a particular case.

DNA Decoding and Fractional Order Analysis. DNA is made up of two polymers forming a double helix and containing four nitrogenous bases, namely, {thymine, cytosine, adenine, guanine}, usually represented as $\{T,C,A,G\}$. Each base on one side bonds with just one type of base on the other side forming the base pairing $A-T$ and $C-G$.

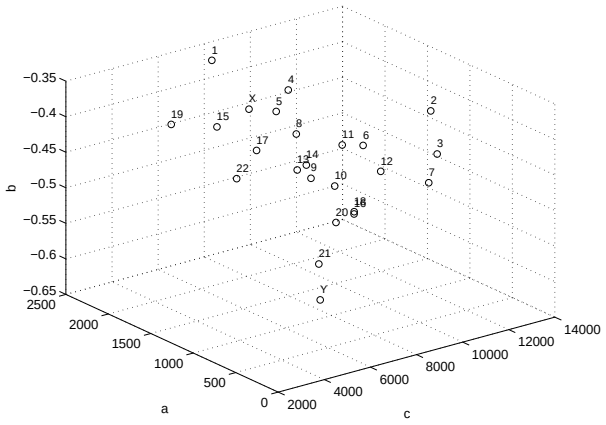
For processing the DNA information we need to convert the four symbols alphabet $\{T,C,A,G\}$ into a numerical value. In fact, the available data includes a fifth symbol, represented by N, which is considered by DNA researchers to have no practical meaning for the decoding and, therefore, N is considered as "zero" during the calculations. In this paper is adopted the direct symbol translation $A=1+i0$, $C=-1+i0$, $T=0+i$, $G=0-i$, $N=0+i0$. Note that the proposed translation follows the base pairing leading to $A=-C$, $T=-G$ and A "orthogonal" to T.

We consider that we move along the DNA strip, one symbol (base) at a time, and that the resulting values form a "signal" $x(t)$ being t denoted loosely as the "time". This signal can be treated by the Fourier transform (FT). In the present case we calculate the FT for 500 sampling frequencies in the interval $10^{-7} \leq \omega \leq 10^{0}$. Since the resulting plot has a significant level of noise, it is applied a median filter to the amplitude of the FT.

After performing several tests it was concluded that a power law using three parameters (PL3), $|F(j\omega)| \approx a\omega^{b}+c$, $a,c>0$, $b<0$, resulted in a good approximation. For adjusting the PL3 was adopted a standard genetic algorithm with elitism, crossover within all population and 5% mutation probability. Furthermore, was adopted a population of 20000 individuals and 5000 iterations of the genetic algorithm.

The parameters $\{a,b,c\}$ of the PL3 approximation constitute a good tool for describing the characteristics of the chromosomes and to map their relationship. In fact, applying the PL3 to the twenty four chromosomes of the Human leads to the three dimensional locus represented in Figure 1. We verify the emergence of a graphical object including an ordering of the chromosomes, with 1 at one end and Y at the other. Rotating and zooming the graphical object we conclude that this locus reveals that some chromosomes have larger similarities (e,g., pair Hu3-Hu4, or pair Hu20-Hu21) while other are considerably distinct (e.g., pair Hu1-HuY, or pair Hu2-Hu20).

[1] Machado J., Kiryakova V., and Mainardi F.: *Recent history of Fractional Calculus*. Communications in Nonlinear Science and Numerical Simulations, 16 (2011), 1140–1153.
[2] Machado J., Costa A., and Quelhas M.: *Fractional dynamics in DNA*. Communications in Nonlinear Science and Numerical Simulations, 16 (2011), 2963–2969.



Locus $\{a,b,c\}$ of the parameters of the PL3 approximations for the set of 24 chromosomes of the Human being