

Topics in Sequence Comparison and Discrete Structures

Christian Houdré and Heinrich Matzinger
School of Mathematics
Georgia Institute of Technology

Final Performance Report: H98230-09-1-0017

Research has been pursued on sequence comparison problems as outlined in the proposal. Progress has been accomplished on topics such as the nature of optimal alignments between sequences or the determination of the order of the variance in a realistic model of DNA. Further progress has been accomplished on other aspects of the project such as the determination of the limiting laws as well as some large deviations principles when the alphabet size and the word length simultaneously grow without bound, or estimates on rates of convergence. Our probabilistic approach has allowed us to completely solve the alternating subsequence case (in mean, variance and limiting law) both for random permutations or random words. This approach has also been extended to a Markovian setting and to pattern avoiding random permutations or words.

The following six manuscripts have been (or will soon be) posted on arXiv or are published and NSA support acknowledged:

Closeness to the diagonal for longest common subsequences (arXiv:0911.2031)

On the rate of approximation in finite-alphabet longest increasing subsequence problems (arXiv:0911.4917)

A Probabilistic Approach to the Asymptotics of the Length of the Longest Alternating Subsequence (The Electronic Journal of Combinatorics, Vol. 17, 2010)

Asymptotics for random Young diagrams when the word length and alphabet size simultaneously grow to infinity (Bernoulli, Vol.16, 2010, 471-492)

Large deviations for the shape of random RSK Young diagrams when the word length and alphabet size simultaneously grow to infinity

Sparse long blocks and longest common subsequence

Talks on the topics of the sponsored research have been given at various levels (undergraduate seminar, graduate seminar, research seminars) and the results disseminated to a large audience.