

1079-92-393

**Angela Angeleska\***, aangeleska@ut.edu, and **Sabrina Kleessen** and **Zoran Nikoloski**. *Integer Quadratic Programming and the Sequence Reconstruction Problem*.

Assembly of genomes from high-throughput data generated by the Next Generation Sequencing (NGS) technologies remains one of the most challenging tasks in modern biology. Here we address the Sequence Reconstruction (SR) problem whereby, for a given collection of subsequences or factors, one is to determine the set of sequences compliant with the collection. First, we give an overview of the SR problem from a language-theoretic perspective, and present the advantages and shortcomings of the existing algorithmic approaches. We then propose an optimization-based formulation which casts the problem as an integer quadratic program whose solutions can be enumerated with appropriate transformations. Finally, we consider three different variants of the SR problem with respect to the input: (1) subset of factors, (2) subset of paired factors, and (3) subset of factors and their inverses. Each of these variants is relevant for the genome assembly problem and can be matched with the employed NGS technologies. The modification of the constraints imposed in the formulation of the respective programs and the different means for their relaxation are also discussed. (Received January 18, 2012)