

# SYSTEMS OPTIMIZATION

Our research and teaching activities center around optimization problems that arise in the context of the analysis and the design of complex systems. On the one hand, we focus on the theory and the development of randomized search algorithms, in particular biologically inspired techniques. On the other hand, these methods are used to tackle cutting-edge optimization problems in Information Technology as well as Molecular Biology.

As to Information Technology, the growing complexity, heterogeneity, distribution, and integration of interconnected systems poses a major challenge. In particular, the optimization of such systems is demanding – due to, e.g., multiple conflicting objectives such as performance and cost, the large number of design choices and the resulting combinatorial explosion of the search space, and uncertain and noisy data. Similar issues arise in Molecular Biology: here, models of specific physiological processes need to be optimized with respect to various types of measurements. However, the vast amount of data produced by high-throughput technology render novel computational tools mandatory in order to analyze the experimental results and to extract useful information.

As a consequence, in both Information Technology and Biology, major advances require corresponding optimization techniques, and in recent years methods that are based on models of biological processes such as evolution and population dynamics have become increasingly important. They possess several desirable features and have been successfully used in numerous application areas where the complexity of the underlying problems prevents exact, traditional algorithms from being applicable.

Our main goal is to improve our ability to tackle the rapidly increasing complexity of today's and future optimization problems by a combination of theoretical, method-oriented, application-driven research. In the following, this will be illustrated on the basis of selected research topics.

## Theory: Analysis of Randomized Search Algorithms

Evolutionary algorithms are probabilistic search techniques that are inspired by models of natural evolution. The underlying principles are simple, but nevertheless these algorithms usually exhibit a complex behavior which is hard to analyze theoretically. Consequently, a lot of empirical knowledge and successful applications have been reported in the literature, but much less rigorous theoretical results about their efficiency are available. Besides empirical investigations, though, theoretical work is important for making more general, but also more precise statements about the performance of evolutionary strategies and for better understanding the underlying dynamics.

Our research in this area encompasses both convergence and running-time analyses. We have studied the behavior of several randomized search algorithms for various scenarios involving multiple optimization criteria and have been able to validate empirical results from previous studies. The insights gained from these investigations have lead to novel concepts and stimulated several algorithmic improvements.

## Methods: Multiobjective Search Space Exploration

Almost every real-world problem involves simultaneous optimization of several incommensurable and often competing objectives. While in the presence of a single objective the optimal solution is usually clearly defined, this does not hold for multiobjective optimization problems. Instead of a single optimum, there is a set of alternative trade-offs, generally known as Pareto-optimal solutions. These solutions are optimal in the wider sense that no other solutions in the search space are superior to them when all objectives are considered. Identifying or approximating the set of Pareto-optimal solutions is beneficial in many application domains as it helps to choose an appropriate compromise solution.

Since generating the Pareto-optimal set is often infeasible for large-scale problems, we have developed state-of-the-art evolutionary algorithms that provide an approximation of the trade-off

surface, see Fig. 1. The strengths of these methods has been demonstrated on various applications, and they are in use worldwide. Furthermore, we have worked on a plug-and-play programming interface, named PISA, for randomized search algorithms in general. This interface enables the exchange of pre-compiled implementations of both optimization methods and benchmark problems without the need of additional programming efforts; the PISA website offers various modules for free download as shown in Fig. 2.

## Applications: Analysis of Biological Networks

With the advent of novel experimental technologies, large amounts of quantitative biological data have become available that allow to study cells on a systems level; nowadays, it is possible to measure expression levels of thousands of genes in a single experiment and to determine protein interactions and localizations on a proteome-wide scale, to name only a few examples. The major challenge in this context consists in revealing structure and functionality of the superimposed networks, which – within single cells – range from metabolism to protein interaction and gene regulation, on the basis of experimental data. To this end, computational and statistical data analysis tools are required that on the one hand provide

decision support to discriminate between different plausible network models, and on the other hand, guide the design of new experiments.

In the context of the interdisciplinary ETH project “Reverse Engineering of Genetic Regulatory Networks”, we have developed advanced optimization methods for identifying groups of co-regulated genes on the basis of mRNA profiling measurements. Our contributions include general, exact as well as specialized, randomized biclustering algorithms that are capable of finding local patterns in large data matrices. The corresponding methods have been integrated within a general-purpose clustering toolbox, cf. Fig. 3, and applied to several data sets from different organisms. Fig. 4 provides an example of a cluster found by one of the algorithms; although multiple time course data sets from different laboratories were combined, the algorithm was able to identify a group of genes that is highly significant regarding common regulation motifs.

<http://www.tik.ee.ethz.ch/sop/>



Eckart Zitzler, born 1970, has been Assistant Professor for Systems Optimization since August 2003. He studied in Oldenburg, Dortmund, and Bochum, Germany, and received the diploma degree in Computer Science from the University of Dortmund in 1996. He then pursued a Ph.D. at ETH Zürich (1996-1999), where he obtained the Dr. sc. techn. degree and simultaneously the high school teaching diploma in 2000. Afterwards, he worked as a senior research assistant at the Computer Engineering and Networks Laboratory at ETH Zürich, interrupted by a research stay at the Institute for Systems Biology, Seattle WA, in 2002. His research focuses on the development and investigation of black-box optimization methods, in particular biologically inspired techniques, and their application to complex optimization problems in the area of Computer Engineering and Molecular Biology. He is one of the initiators of the biennial International Conference on Evolutionary Multi-Criterion Optimization and was General Co-Chair in 2001, 2003, and 2005.

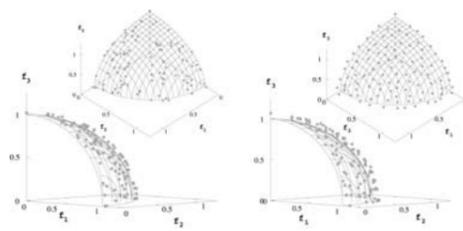


Figure 1: Trade-off solutions obtained by two different algorithms for an optimization problem with three objectives.



Figure 2: The PISA website provides implementations for various optimization problems and algorithms for different platforms.

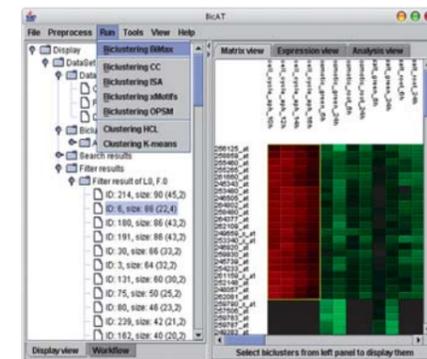


Figure 3: Snapshot of the biclustering toolbox BicAT developed in our group; on the right, one of the biclusters found is visualized.

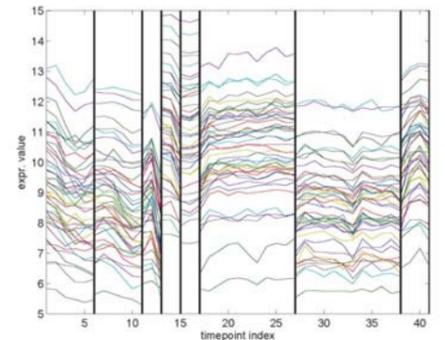


Figure 4: Example of a gene cluster over multiple, heterogeneous time course experiments