1 Theory and Methods

The research block *Theory and Methods* is divided into four work packages given below.

WP1 – Optimization

Sub-coordinator: Prof. Westerlund. In the area of optimization, the focus will be on mixed integer global optimization, model representations and generic methods and techniques for global optimization of certain classes of large scale mixed-integer nonlinear problems. Mixed-integer nonlinear model reformulation and transformation techniques, model representations using nonsmooth assumptions, distributed computing methods and computational tests on different large scale applications in global optimization, such as quadratic assignment, are subtopics within this area. A specific goal within the COE is the creation of theory and algorithms for nonsmooth cutting-plane techiques. The work within WP1 will be closely linked to all the applied WPs.

WP2 – Systems engineering

Sub-coordinator: Prof. Toivonen. In systems engineering, one focus will be on solving complex control problems, for which traditional controller design methods are inadequate. These problems include feedback control of systems which involve integer-valued variables, nonlinear systems and multiple objective controller design problems. The only feasible way to solve these problems is by numerical optimization, and the research on optimization techniques (*cf.* WP1) will therefore be directly applicable to these problems.

A second focus area will be to develop system identification techniques, where models of dynamical systems are inferred from measurements obtained from the system variables. Although system identification is an old research topic, it is still a very active and important field propelled by new theoretical advances and challenging applications. One recent focus area of the systems engineering group is robust support vector methods for nonlinear system identification, *cf.* [4]. In the planned project, these methods will be developed further and new challenging problems will be addressed, such as identification of systems which switch between a set of different behavioral modes at *a priori* unknown time instants. Identification of such systems is important for example in modeling gene regulatory networks, where the properties are determined by the set active genes (*cf.* WP6, task 2).

A third research area in this work package is the modeling and control of periodic systems, which is a very timely topic with important applications in active vibration control in electrical machines and generators (*e.g.*, in wind turbines), *cf.* WP7, task 2. In the planned work, basic problems in representing time-periodic systems for identification will be addressed, where mathematical systems theory (*cf.* WP3) will play an important role.

WP3 – Systems theory

Sub-coordinator: Prof. Staffans. Many of the processes that one tries to control are made up of an infinite number of distributed parameters, and they are consequently most appropriately modeled as infinite-dimensional systems (such as partial differential equations or delay equations). A natural setting in the case of a continuous time

variable is to view them as infinite-dimensional well-posed linear systems. This theory is under active development within the COE. Another approach that has recently emerged is to replace the classical input/state/output point of view by a "state/signal" point of view, where inputs and outputs have been combined into one common signal describing the interaction between the system and the surrounding world in the tradition of classical circuit theory. One goal is to further develop various aspects of this theory.

WP4 – Mathematical statistics

Sub-coordinator: Prof. Corander. Very recently, it has been claimed that the Bayesian paradigm has revolutionized statistical thinking in numerous fields of research, as a considerable amount of novel Bayesian statistical models and estimation algorithms have gained popularity among scientists. Despite of the evident success of the Bayesian approach, there are also many research problems where the computational challenges have so far proven to be too exhaustive to promote wide-spread use of the state-of-the-art Bayesian methodology. In particular, due to significant advances in measurement technologies, *e.g.*, in molecular biology, a constant need for analyzing and modeling very large and complex data sets has emerged on a wide scale during the past decade.

The prevailing situation is somewhat paradoxical, as the theoretical superiority of the Bayesian paradigm as an uncertainty handling framework is widely acknowledged, yet it can be unable to provide practically applicable solutions to complex statistical problems. To resolve this issue, the research within WP4 will have a focus on stochastic computational and modeling strategies to develop algorithms that overcome problems associated with the analysis of highly complex data sets. With these methods we aim to be able to solve a multitude of statistical learning problems for data sets which cannot yet be reliably handled in practice by any of the existing Bayesian tools. Our approaches will build upon recent advances in Bayesian predictive modeling and adaptive stochastic Monte Carlo computation, to create a novel family of parallel interacting learning algorithms.

2 Applications in science and engineering

The research block *Applications in science and engineering* is divided into three WPs.

WP5 – Physics and materials science

Sub-coordinator: Dr. Pörn.

Task 1 – Minimizing the energy of electron configuration in disordered materials. Prof. Westerlund in cooperation with Prof. Ronald Österbacka and Fredrik Jansson, Department of Physics, ÅAU. The Coulomb glass is a model for a disordered material, where the conduction electrons are localized to impurity sites and the electrons strongly interact with each other. This can be realized, for example, in lightly doped semiconductors at a low temperature. The sites are assumed to be randomly distributed in space. Typical for glassy systems is that their dynamics are very slow. If the system is cooled rapidly it takes a very long time for the electrons to reach equilibrium. [1, 2, 3]

We assume that the system consists of N sites, and that n of these are occupied by electrons. Assuming that the sites are neutral when empty, the total energy of the system can be written as

$$E = \frac{e^2}{4\pi\varepsilon}\ \frac{1}{2}\sum_{i}^{N}\sum_{j\neq i}^{N}\frac{x_i x_j}{r_{ij}} + \sum_{i}^{N}\epsilon_i x_i \ \equiv\ \boldsymbol{x}^T\boldsymbol{M}\boldsymbol{x} + \boldsymbol{c}^T\boldsymbol{x}, \tag{1}$$

where r_{ij} is the distance between sites i and j, ϵ_i is a random energy specific for the site i and $x_i \in \{0, 1\}$ is the occupation number of site i. This is illustrated in figure 1. If n sites are occupied we have $\sum_i^N x_i = n$.

The ground state of the system is defined as the electron configuration that has the smallest total energy. Minimizing E, with the constraints that all x_i are binary variables, and that exactly n of them are equal to 1, forms a binary quadratic optimization problem. So far, the ground state has been searched for by various simulated annealing techniques. These typically give good results, in the sense that they quickly find electron configurations with low energies, but the methods can not prove that the minimum has been found and do not provide any bounds for it. By treating the problem as a quadratic assignment problem, one can prove that the solution is indeed optimal

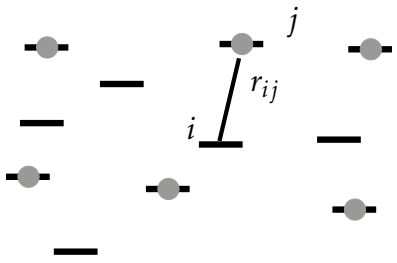

Figure 1: Illustration of the distance between two sites i and j of conduction electrons

or at least obtain upper or lower bounds for the minimum. A remaining challenge is the size of the system. To get physically reasonable results the system should be large, preferably containing several thousands of sites. Systems of this size are very difficult to optimize. The problem is closely linked to the research done in WP1.

References

[1] Shklovskii, B.I., Efros, A.L., Electron properties of doped semiconductors, *Springer-Verlag*, Berlin, 1984.

[2] Ortuño, M., Caravaca, M. and Somoza, A.M., Numerical Study of Relaxation in Coulomb Glasses, Physica Status Solidi (c), **5**, 674–679, 2008.

[3] Tsigankov, D.N., Pazy, E., Laikhtman, B.D and Efros A.L., Long-time relaxation of interacting electrons in the regime of hopping conduction, Physical Review B, **68**, 184205, 2003.

[4] Tötterman S. and Toivonen H.T., Support vector method for identification of Wiener models, Journal of Process Control, **19**, 1174–1181, 2009.

WP 6 – Biotechnology and -medicine

(Sub-coordinator: Doc. Petterson and Andreas Lundell).

Task 1 – Peptide docking. Andreas Lundell in cooperation with Prof. Christodoulos Floudas, Princeton University. In peptide docking problems, equilibrium structures are identified for macromolecule-ligand complexes. Additionally, equilibrium structures are identified for a number of candidate "docking" molecules and a measure of their relative binding affinity is quantified and compared. The ability to predict these structures is of great theoretical interest and will greatly increase the understanding of hereditary and infectious disease and aid the interpretation of genome data.

A part of the peptide docking problem can be interpreted as a global optimization problem, where the total conformation of energy, describing the molecules in terms of atomic bounds and effective interactions, is minimized. In the ECEPP/3 (empirical conformational energy program for peptides) potential model considered, the contributions to the total conformation energy are the electrostatic, nonbounded, hydrogen bounded, torsional, cystine loop-closing and cystine torsional energies. The energy minimization problem is a highly nonconvex optimization problem describing the peptide chain. To solve problems of this type, the current αECP algorithm and the algorithm for signomial functions [3] will be used and extended. The work within this task will be closely linked to WP1.

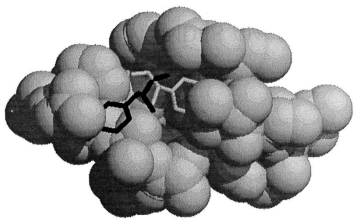

Figure 2: Local (black) vs global (white) minimum configurations of tyrosine

In figure 2, a local and a global minimum energy conformation of tyrosine from [1] are shown. Observe that the structure of the solutions are quite different, although the conformation energies are very close. The energy values corresponds to -823.3 kJ/mol and -833.0 kJ/mol respectively in the illustrated case.

References

[1] Floudas, C.A., Deterministic Global Optimization – Theory, Methods and Applications, Kluwer Academic Publishers, 2000.

[2] Floudas, C.A., Klepeis J.L. and Pardalos P.M., Global Optimization Approaches in Protein Folding and Peptide Docking, DIMACS Series in Discrete Mathematics and Theoretical Computer Science, (Ed. F. Roberts), **47**, 1999.

[3] Lundell A., Westerlund J. and Westerlund T., Some transformation techniques with applications in global optimization, Journal of Global Optimization, **43**, 391-Ű405, 2009.

Task 2 – Systems biology. Doc. Petterson partly in cooperation with Prof. Eriksson, Department of Biology. Mathematical modeling has become increasingly important in the study of complex dynamical cellular processes, such as gene regulatory networks. Cellular processes consist of biochemical reaction networks, which can for example by described by nonlinear differential equations. In practice, the reaction rate constants are unknown and must be identified from concentration measurements. The identification problem is, however, complicated by the fact that only a few of the species concentrations can be measured. Moreover, the concentration measurements typically represent average values for a cell population rather than the concentration of a single cell. The purpose is to develop rigorous methods for the identification of rate constant and unmeasured concentrations and their intercellular distributions for such systems. The methods will be applied in joint work with researchers in cell biology to develop mathematical models of complex cellular processes.

Task 3 – Statistical methods in medicine. Prof. Corander. Medical statistics and informatics have in the recent years gone through rapid developments where the need of efficient complex system analysis and modeling tools has become more urgent as the amount of medical information has increased considerably. In particular, the technological advances in genomics, proteomics and expression platforms have generated vast data masses and the challenge remains on how to transform these data into knowledge that improves our understanding of disease aetiology and potential in candidate therapeutic actions.

In many medical contexts it is necessary to use unsupervised statistical learning methods to discover mechanisms related to the evolution of diseases from various biological databases. We have recently illustrated the potential of such an approach in the context of antibiotic resistance development in the pneumococcus [2] and in genomic large-scale cancer profiling [3]. The methods developed within Prof. Corander's research group have considerable potential for further diverse medical applications, however, the computational algorithms must still evolve to meet the challenges stated by the next generation of molecular biological information that will be available

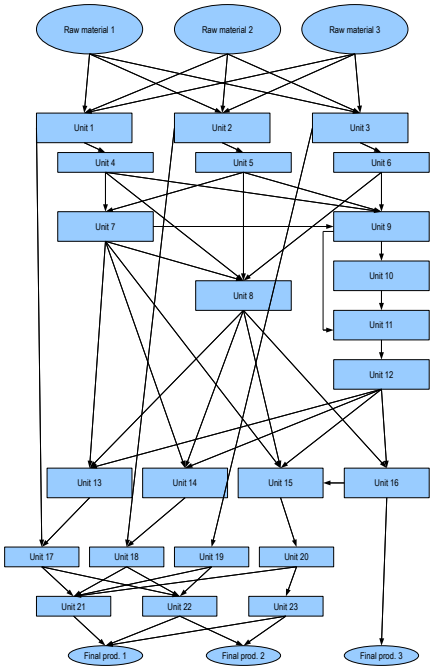

Figure 3: A process flow sheet for a flexible manufacturing network

in the immediate future. Currently, the only potential approach to handle such complex and massive data sets will be to develop the algorithms for a massively parallel computing architecture. Some steps towards stochastic statistical learning algorithms in such a setting were taken by [1], however, considerable theoretical and applied research work is still needed to make the potential of such an approach fully utilizable.

References

[1] Corander, J., Gyllenberg, M. and Koski, T., Bayesian model learning based on a parallel MCMC strategy. Statistics and Computing, **16**, 355–362, 2006.

[2] Hanage, W.P., Fraser, C., Tang, J., Connor, T., Corander, J., Hyper-recombination, diversity and antibiotic resistance in the pneumococcus. Science, **324**, 1454–1457, 2009.

[3] Marttinen, P., Myllykangas, S. and Corander, J., Bayesian clustering and feature selection for cancer tissue samples. BMC Bioinformatics, 10:90, 2009.

WP 7 – Engineering

Sub-coordinators: Doc. Harjunkoski and Doc. Westerlund.

Task 1 – Large scale industrial scheduling. In cooperation with Prof. Grossmann, Carnegie Mellon University. In this task, complex production network problems will be considered. Examples of problems of this kind are industrial production scheduling in production facilities with multiple units using flexible production phases as well as having multiple raw material feeds, multiple product and by-product streams [2, 3].

We will especially consider complex production scheduling networks in the manufacturing industry, where flexible multi-stage production processes are used. The processes considered are considered flexible as a particular product can be manufactured using one of several possible routes. Furthermore, the plant includes several instances of processing units that can be run either in parallel or stand-alone, as shown in figure 3. While the flexibility creates opportunities to greatly improve efficiency using optimal production plans, it inevitably increases the scale of the scheduling problem at hand. The flexibility to choose routes and unit setups introduces bilinear constraints to the mathematical model making it a mixed integer nonlinear programming (MINLP) problem. Furthermore, the large number of raw materials and products, to be scheduled and rescheduled over weekly and monthly horizons, introduces thousands of variables and constraints which results in models of large scales. Different modeling paradigms and solution techniques for the production network is, therefore, focused on. Reformulation techniques in the modeling framework will be considered as an integrated part of possible solution approaches, as effective reformulations are key factors in transforming previously unsolvable problems to a simpler solvable form. The solution methods will include deterministic and stochastic as well as hybrid optimization methods. Thus, the work within this task will be closely interlinked with the methodological work in WP1 and WP4.

Task 2 – Time-periodic systems Prof. Toivonen and Dr. Emet. The control of periodic systems has recently gained wide interest due to important applications in active vibration damping and speed control of periodic rotating machinery, such as internal combustion engines or electrical machines and generators. The systems engineering group has recently developed active cylinder balancing methods for diesel engines, where the cylinder-wise torque contributions are equalized by online minimization of relevant torsion vibration frequency components, *cf.* [4]. Such methods are essential for vibration reduction for example in marine diesel engine installations. In future work, the methods will be developed further to include adaptive algorithms for unknown engine dynamics and to be able to address gas engine with large numbers of cylinders.

Further work will also focus on control of periodic electrical machines and generators, where active vibration control is extremely important for improving operability and reducing wear of equipment. These systems are challenging to control because, in addition to being periodic, the system behavior depends on load and speed. Therefore, new robust control algorithms which respond quickly to disturbances need to be developed for these systems. The theory of system identification and control of time-periodic systems developed in WP2 will be used here.

Advantageous properties of time-periodic systems are utilized in several other engineering systems as well; examples are high -purity separation systems. The separation of the compound betaine is, for example, conducted by differential migration during passage through porous mediums in chromatographic columns under periodic operation. Betaine is a valuable product used for many purposes in the food and pharmaceutical industries. A separation system configuration contains several columns. A high-purity product is obtained by altering the feed stream (molasses), the eluent (wa-

ter) and recycled unpure product streams periodically to the columns. The columns can be modeled by partial differential equations (PDE) to obtain the concentration profiles as functions of time and space. Since the system is operated in a periodic mode the system of PDEs need to be solved as a two-point boundary value problem, where the concentration profiles inside the columns are the same in the beginning and end of the time periods.

The concentrations $c_{kj}(t,z)$ in time and space (column height) can be obtained by solving the following boundary value problem

$$(1+F\beta_j)\frac{\partial c_{kj}}{\partial t} + F\cdot\sum_{l=1}^{C}\beta_{jl}\left(c_{kl}\frac{\partial c_{kj}}{\partial t} + c_{kj}\frac{\partial c_{kl}}{\partial t}\right) + u\frac{\partial c_{kj}}{\partial z} = D_j\frac{\partial^2 c_{kj}}{\partial z^2},$$

$$c_{kj}(t,0) = y_k^{\text{in}}(t)\cdot c_j^{\text{in}} + \sum_{l=1}^{K} x_{lk}(t)\cdot c_{lj}(t,z_H),$$

$$c_{kj}(0,z) = c_{kj}(\tau,z).$$

The index k stands for the column and index j for the component, C is the number of components, K the number of columns, z_H the height of the columns, F the phase ratio, D dispersion coefficients, u the phase velocity, τ the duration of the cycle and β are mass-transfer coefficients. As the operation is affected by a number of discrete decisions, described in the logical functions y_k^{in}, the operational problem becomes a very complex mixed integer nonlinear optimization problem including both integral and differential equation constraints in addition to algebraic constraints [1]. The problem is very demanding and will be studied in this WP, as well as in WP1–WP3.

References

[1] Emet S. and Westerlund T., Solving a Dynamic Separation Problem using MINLP Techniques, Applied Numerical Mathematics, **58**, 395–406, 2008.

[2] Grossmann, I.E., Enterprise-wide Optimization: A New Frontier in Process Systems Engineering, AIChE Journal, **51**, 1846–1857, 2005.

[3] You, F. and I.E. Grossmann, Mixed-Integer Nonlinear Programming Models and Algorithms for Large-Scale Supply Chain Design with Stochastic Inventory Management, I&EC Research **47**, 7802–7817, 2008.

[4] Östman F. and Toivonen H.T., Model-based torsional vibration control of internal combustion engines. IET Control Theory & Applications, **2**, 1024–1032, 2008.