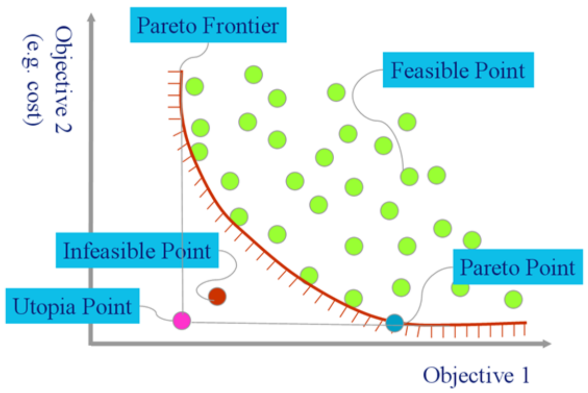
# Research/Master Projects 2014

M. Emmerich

https://moda.liacs.nl

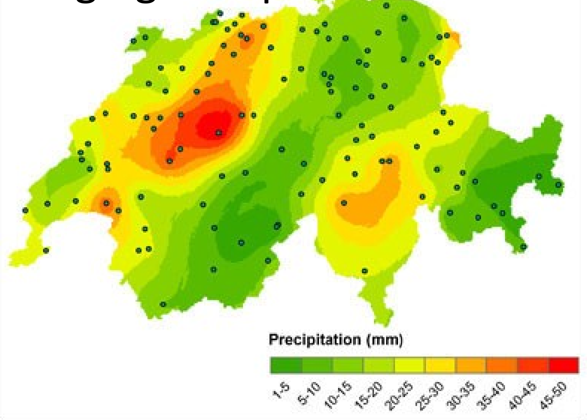
# Multicriteria Optimization, Design, and Analytics (MODA) - Research Group

- We design algorithms for solving optimization and learning problems with multiple constraints and objective functions
- Core technologies:
  - Pareto front approximation with stochastic and deterministic algorithms
  - Modelling of response surfaces (learning functional mappings)
  - Set-oriented approaches to solving classification and prediction problems
- Where used: Scientific Databases, Computer Aided Design, Scheduling

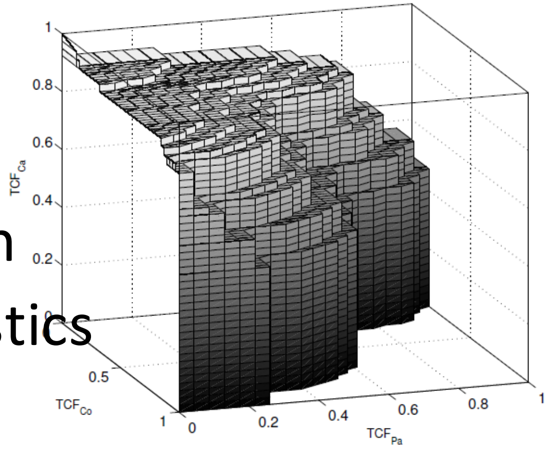



Pareto front (=optimal set) of a problem with 2 objectives.

Kriging interpolation model




3-D ROC of a 3-class classifier




- Core application fields:
  - Drug Discovery
  - Computer Aided Design
  - SPAM/Medical Diagnostics
  - Logistics
  - Engineering Design

# Industry Projects

Projects in close collaboration with industry

# Master‘s Thesis:

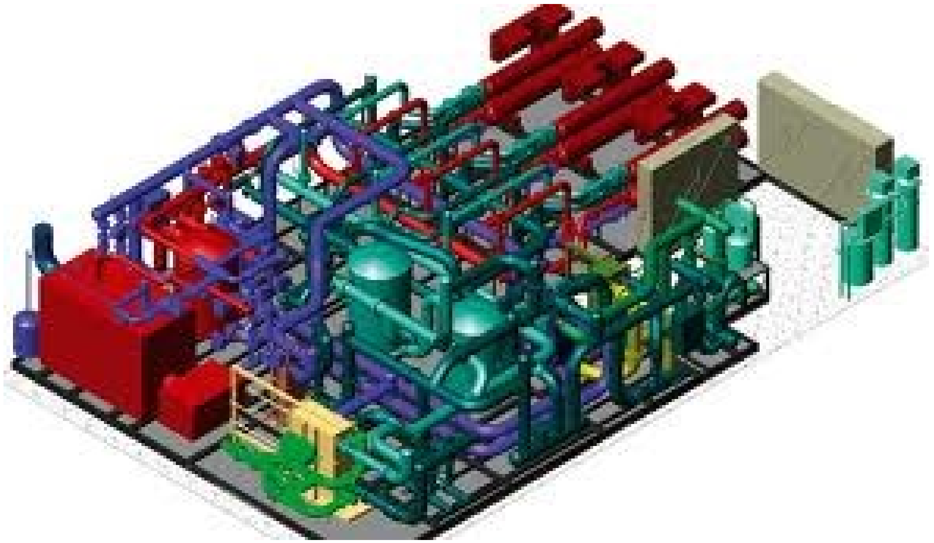
| | |
|---|---|
| Supervisor LIACS: | Prof. dr. Michael Emmerich |
| In cooperation with: | Gijs Willem Sloof (Stabiplan) |
| Start date: | 2.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |

Gijs Willem Sloof MSc managing director Stabiplan BV www.stabiplan.nl disclaimer: http://www.stabiplan.nl/disclaimer

**Design of optimal routing algorithms for distribution systems.**

**A major part of designing the technical installations of a building is the design and coordination of the different distribution systems. In a building project, a lot of work could be saved when the design of these systems can be done more automated. One of the first steps in the design is the routing of the pipes/ducts from the startpoint of the system (e.g. an air handling unit or a boiler) to the endpoints of the system (e.g. air terminals or radiators). For this, the architectural model provides the most important input: where are the rooms, the corridors, what's the available space above the ceiling, etc. Apart from that, there are several practical rules that make one solution more optimal then others. The goal is to design intelligent routing algorithms giving options to the designer for the routing of the systems. The algorithm should be able to generate the routing with minimal user input On the other hand, it must be possible for the designer to specify/change the priority of design criteria, so he can compare different routing alternatives. The routing algorithm should also include the coördination of the different systems so collisions between systems are prevented.**

**An extension of the algorithm is to involve the schematical designs (the installation concept) in the routing algorithm. Then we do not only generate the piping layout, but also the devices and equipment of the distribution system is automatically generated, according to the schematic drawings.**



# Master‘s Thesis:

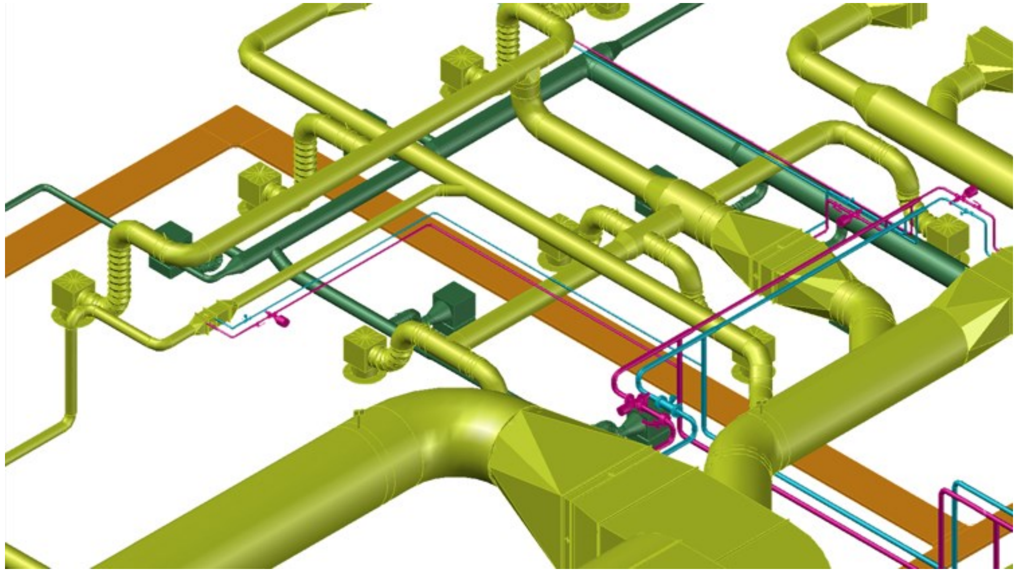
| | |
|---|---|
| Supervisor LIACS: | Prof. dr. Michael Emmerich |
| In cooperation with: | Gijs Willm Sloof (Stabiplan) |
| Start date: | 2.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |

Gijs Willem Sloof MSc managing director Stabiplan BV www.stabiplan.nl disclaimer: http://www.stabiplan.nl/disclaimer

**Design of a generic algorithm for dimensioning of distribution systems.**

**In the field of building services, we have several types of distribution systems, e.g. gas, tapwater, waste water, ventilation ducts, heating/cooling, sprinkler. For each of these systems, calculation algorithms are describe in standard documents (often country specific) how to dimension them. In Stabicad, we've implemented these calculation algorithms as separate calculation modules, with some parts of the implementation shared. When we consider it a bit more abstract, all these calculation algorithms are similar: There are some constraints (e.g. velocity and/or pressure loss restrictions), and we have to find an optimal (often with respect to costs) solution satisfying these constraints. These kind of problems is also suited to be solved with iterative algorithms, which are often easy to implement in software. Especially for circular systems, the iterative way saves a lot of complexity.**

**The goal is to design a generic algorithm being able to calculate all those different types of systems, preferably in a iterative way. This will make it possible to support new types of systems or new localizations of a system with minimal effort. The algorithm should be verified with the implementation of new calculation module or re-implementation of an existing one.**



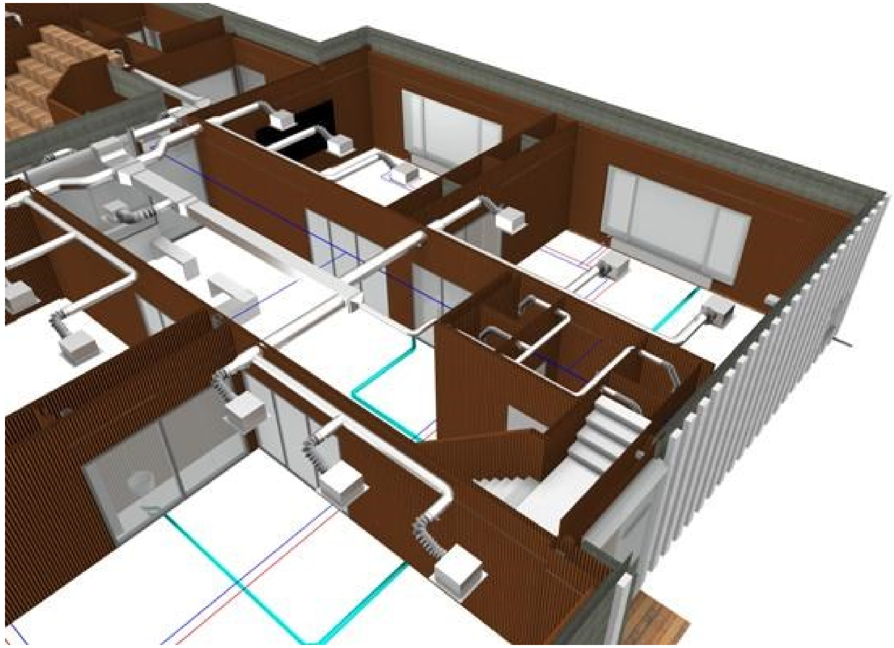
# Master‘s Thesis:

| | |
|---|---|
| Supervisor LIACS: | Prof. dr. Michael Emmerich |
| In cooperation with: | Gijs Willem Sloof (Stabiplan) |
| Start date: | 2.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |

Gijs Willem Sloof MSc managing director Stabiplan BV www.stabiplan.nl disclaimer: http://www.stabiplan.nl/disclaimer

**Generate parametric components from the Stabicad graphic library.**

**Stabicad components that are drawn in a BIM project are generated from our graphic library. The input for this library is a parametric database, fed by manufacturer catalogs, containing size information of the components. The generated components are not parametric, so users cannot change them by adjusting parameter values.. We want to change the way the components are generated: the parameters of the database should be included as adjustable parameters in the components. The main reason is that in a manufacturers catalog, a component is often part of a series, and users want to be able to easily change a component to another instance of the series. So the goal of this project is to implement the parametric structure of the database in the generated components and to design a infrastructure to implement the concept of series in these components.**



# Master's Thesis: An Evolutionary Machine Learning Engine and its use for SPAM and/or Medical classification*

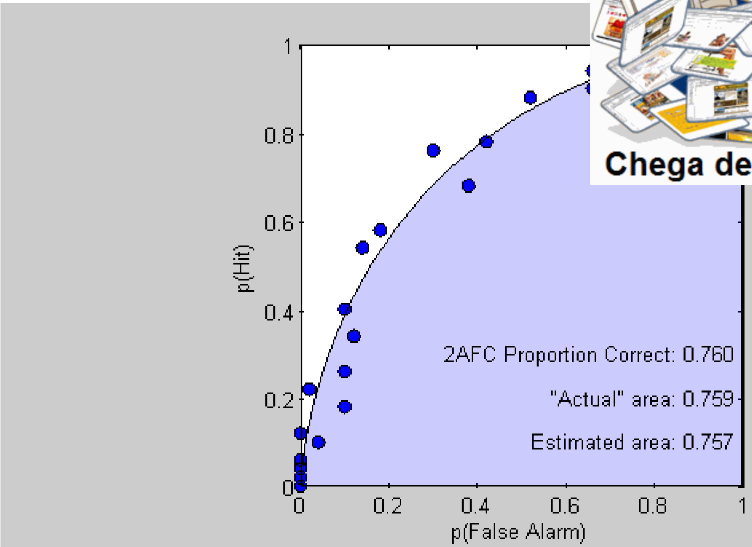
| | |
|---|---|
| Supervisor LIACS: | Prof. dr. Michael Emmerich |
| In cooperation with: | Dr. Vitor Basto Fernandes (ICCI Leiria, Portugal) |
| Start date: | 2.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |

The research is conducted in collaboration with researchers from ICCI Portugal and University Vigo.


**Summary:**

**Researchers from LIACS have recently discovered a new algoroithm for learning classifiers based on RoC curve approximation --- it is based on multiobjective evolutionary optimization and called CHEMOA. The algorithm seeks to find RoC curves with minimal area under the RoC. Currently it maximizes the number of correct alarms while minimizing the number of false alarms. The new research will extend the capability towards minimal parsimony – that is the number of rules for machine learning is to be minimized.**

**Prerequisites (helpful, not mandatory): Interest in data mining, machine learning, and evolutionary algorithms. Source code is available in JAVA and needs to be extended**



# Master‘s Thesis: Buiding Design Optimization in 4-D*

| | |
|---|---|
| Supervisor LIACS: | Prof. dr. Michael Emmerich |
| In cooperation with: | Dr. Zhiwei Yang |
| Start date: | 2.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |

**Summary:**

**The goal is to design buildings based on four objectives -- cost, comfort, energy, and robustness. We have a simulator for buildings that can be used as a black box objective function. The challenge is to find 4-D Pareto fronts and visualize them using transparent plots. Recently there exists a software for transparent plots. The task is to couple this software with a 4-D evolutionary multiobjective optimizer – SMS-EMOA and to test this approach on a real world scenario.**

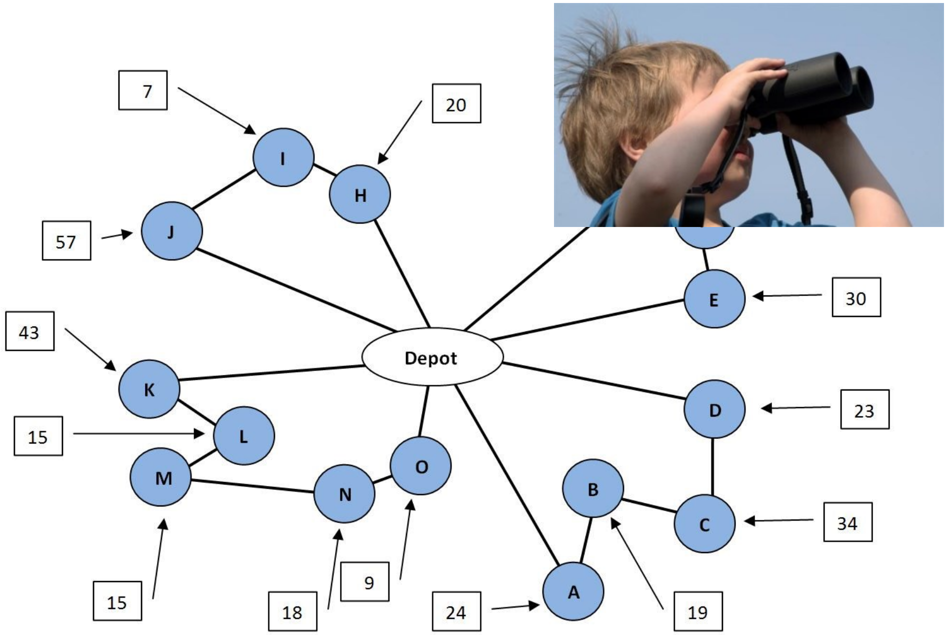
**Prerequisites (helpful, not mandatory): Interest in real world optimization problems, OpenGL, basic knowledge evolutionary algorithms**

# Master Thesis: Multicriteria Network Analytics for Vehicle Routing

| | |
|---|---|
| Supervisor LIACS: | Prof. dr. Michael Emmerich |
| In cooperation with: | Dr. Zhiwei Yang |
| Start date: | 2.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |

In collaboration with Trigion

**Summary: Dynamical vehicle routing makes it possible to plan a fleet of vehicles over a period of time with new events coming in during the day. A challenge is to predict new events based on past distribution of events. An example is given by a emergency company for which alarms in different nodes of the network will be simulated based on the distribution of past events. The task is to learn event distributions in networks and use them in predictive optimization. We will collaborate with a security company (Trigion, NL).**

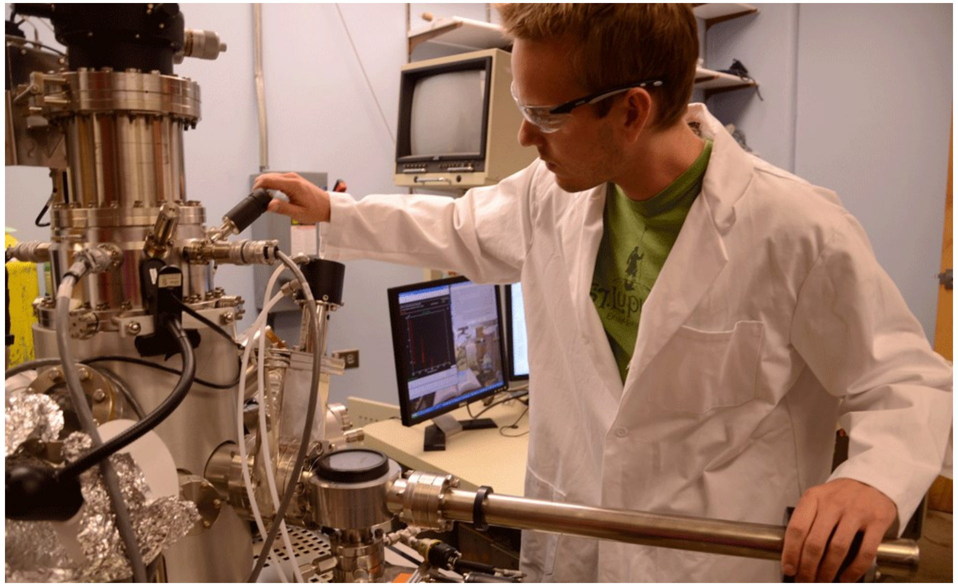



**Prerequisites (helpful, not mandatory):**

**Knowledge in Vehicle Routing or TSP, Programming in JAVA and C++. Programming with threads. Interest in probability and statistics.**

# Multidisciplinary Projects

Applying MODA in other scientific fields



# Master Project: Optimal Selection of Drug Candidates in Molecular Databases

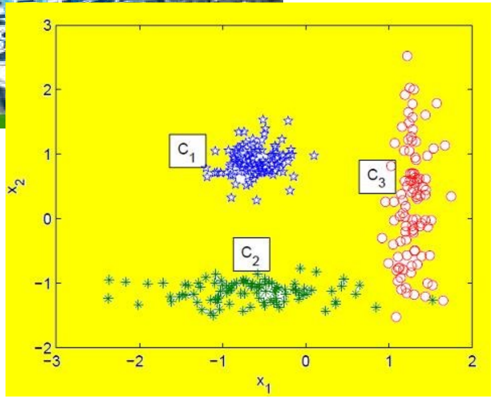
| | |
|---|---|
| Supervisor LIACS: | Prof. dr. Michael Emmerich |
| In cooperation with: | Dr. Iryna Yevseyeva, Bart Lenselink (LACDR) |
| Start date: | 2.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |

The research will be in collaboration with the Leiden Academic Center of Drug Research

**Summary:** In high throughput screening the problem is to select subsets of chemicals from large databases. Portfolio optimization balances risk and return. To find an optimal portfolio a small subset has to be extracted from a large set – an NP hard problem. Instead of quadratic programming, which is too slow for large sets, a fast construction and genetic algorithms will be used and tested on some problem instances.

**Prerequisites (helpful, not mandatory):**

**Interest in Algorithm Design, Statistical Quality Assessment, Interest in Cheminformatics (no specific knowledge in chemistry will be required).**



# Master Project: Genetic Programming for Making Bit Coins?

| | |
|---|---|
| Supervisor LIACS: | Prof. dr. Michael Emmerich |
| In cooperation with: | Prof. Thomas Bäck |
| Start date: | 2.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |

**Summary:**

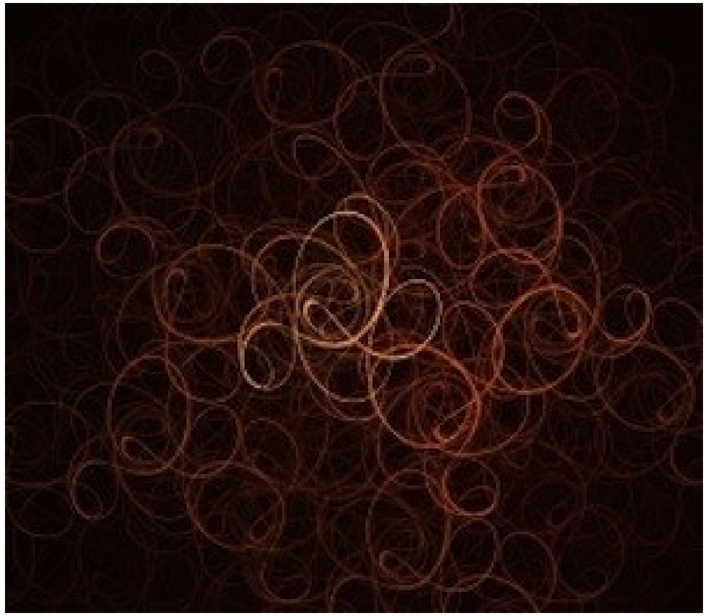
**The bitcoin is a virtual currency based on cryptographic functions. In the SHA-256 cryptography algorithms preimages of a function needs to be computed. BitCoin uses the SHA-256 hash algorithm to generate verifiably "random" numbers in a way that requires a predictable amount of CPU effort. Generating a SHA-256 hash with a value less than the current target solves a block and wins you some coins. This means that the deviation from a result needs to be minimized. Genetic programming are powerful minimization algorithms. The assignment it to represent the problem of finding a preimage of a hash funtion (SHA-256) as a search problem for genetic algorithms and find out how close these algorithms get to preimages. There is a small chance that they are even capable to solve the problem, in which case the bachelor thesis will yield a major result – i.e. an algorithm that creates money.**

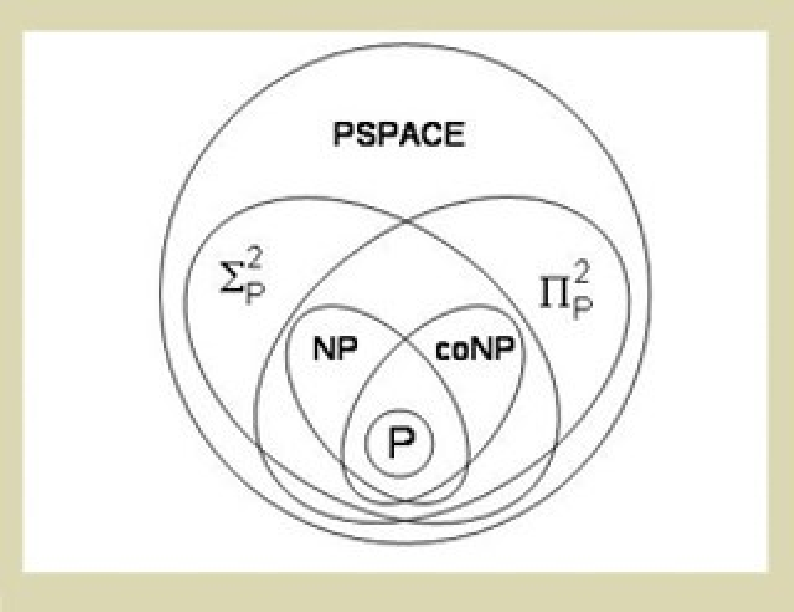
**Prerequisites (helpful, not mandatory):**

Interest in genetic algorithms and cryptography. Curiosity. Programming skills in C.

# Theory and Algorithm Design

Developing theoretical foundations of optimization and new algorithms





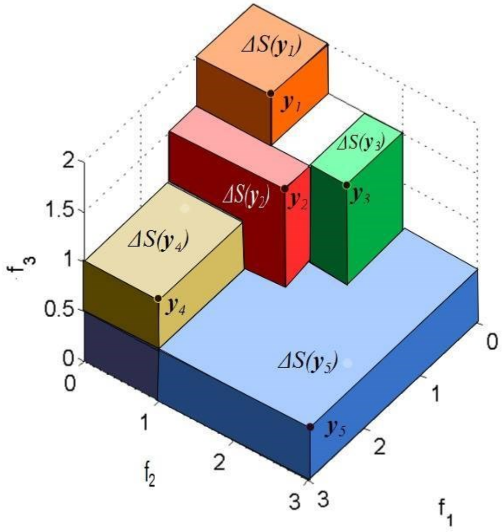
# Research Project: Hypervolume Subset Selection – A candidate for an NP hard problem?

| Supervisor LIACS: | Prof. dr. Michael Emmerich |
|---|---|
| In cooperation with: | Dr. Andre Deutz |
| Start date: | 2.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |

Note:The supervisor is a leading Researcher on this question and has developed HSSP selection for special cases. There are some ideas that can be tried.

**Summary: Prerequisites (helpful, not mandatory):**

Consider problems where n axis aligned boxes cover a volume and a selection of k< n boxes need to be selected that together (union) covers a maximal volume. In 2D dynamic programming can be used, yielding a cubic time complexity, Can it be done faster? It is conjectured that hypervolume subset selection is NP hard in more than two dimensions. Is that so? The assignment will be to work on this difficult question and look for results in the field of paramererized complexity that might point towards solutions. Any progress towards the solution of this problem is very valuable to the advancement of multiobjective optimization methods.




- Interest in computational geometry and complexity

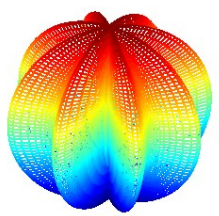
# Research Project: Diversity Optimization

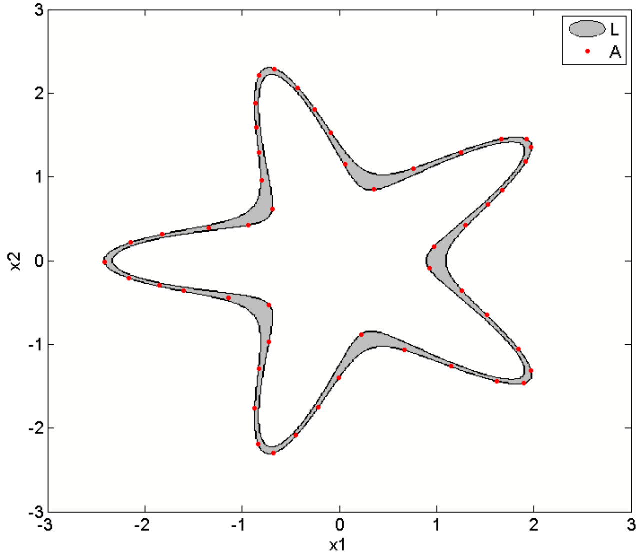
| Supervisor LIACS: | Prof. dr. Michael Emmerich |
|---|---|
| In cooperation with: | Dr. Iryna Yevseyeva |
| Start date: | 2.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |

**Summary: Prerequisites (helpful, not mandatory):**

Finding a diverse set with prescribed properties can be formulated as an optimization problem. Since long diversification plays an important role in minimizing risks for portfolios. But there is also a great potential in using diversification in optimization, in particular in constraint optimization and in cases where objectives give rise to alternative solutions that all require inspection. More recently Diversity has been put on a axiomatic basis and this has given rise to systematic algorithm design approaches for diversity optimization. First prototypes of such algorithms have been developed in LIACS and applied for drug-discovery. The assignment is to experimentally investigate these, for instance by approximating natural shapes represented in Gielies superformula.

Prerequisites: - Interest and knowledge in evolutionary algorithms, computer graphics





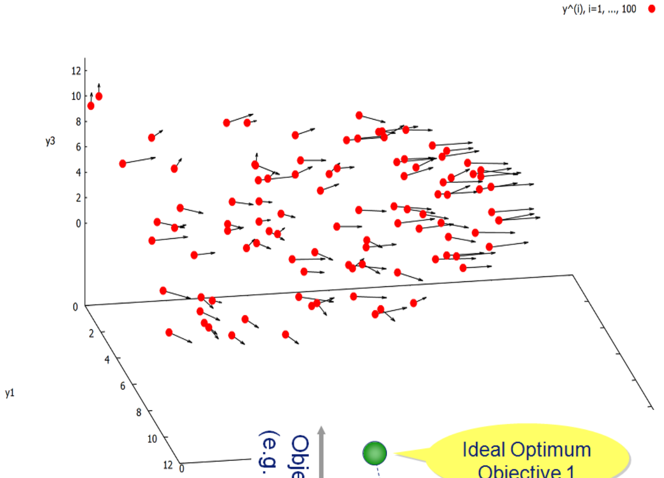
# Master Project: 3-D Newton‘s method for Pareto optimal sets

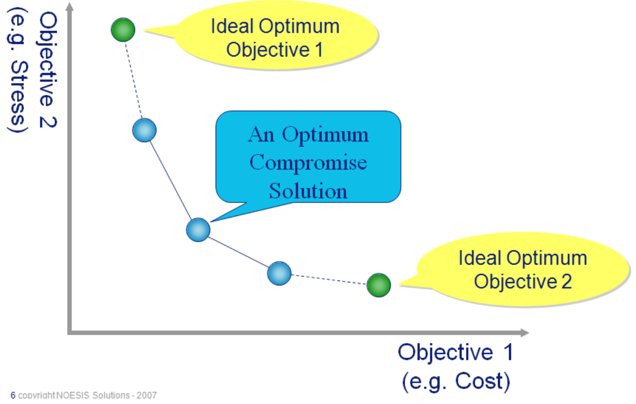
| | |
|---|---|
| Supervisor LIACS: | Prof. dr. Michael Emmerich |
| In cooperation with: | Dr. André Deutz |
| Start date: | 2.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |

In colaboration with Dr. Oliver Schütze, CINVESTAV Mexico

**Summary:**

**The hypervolume gradient is a field of vectors that point in the direction of maximal net-improvement of a Pareto front approximation in multiobjective optimization. The structure and computation of it has been described in 2014 by André Deutz and Michael Emmerich from Leiden university. It is expected that using the gradient information will significantly enhance continuous multiobjective optimization methods. The aim of the thesis is to develop gradient based algorithms for multiobjective optimization and test their performance on the state-of-the art set benchmarks, both in 2-D and 3-D.**



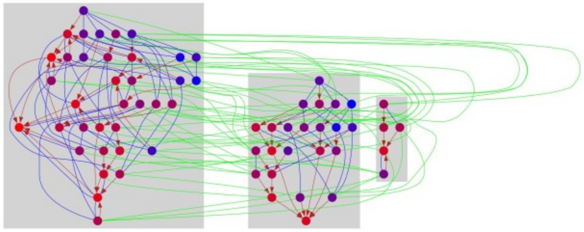

**Prerequisites (helpful, not mandatory):**

Interest in optimization algorithms and numerics.

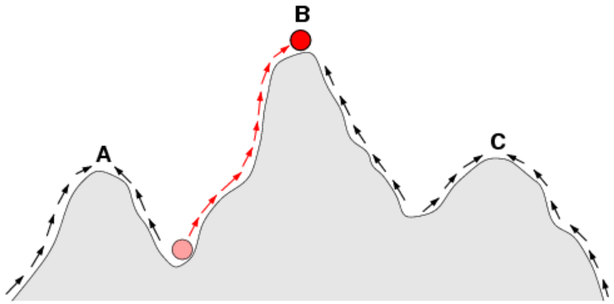
# Research Project: Local Optima Networks in Discrete Optimization

| | |
|---|---|
| Supervisor LIACS: | Prof. dr. Michael Emmerich |
| In cooperation with: | Prof. Thomas Bäck |
| Start date: | 1.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |

Note: The project is in collaboration with Almende e.V., Rotterdam, who provides professional solutions for VRP problems.



**Summary:** Vehicle Routing Problems and Nk Landscapes are examples for combinatorial optimization problems. Local Search and Evolutionary Optimization are often used to find global optima for these problems. Though, gradual improvement strategies are often proposed, it is widely unclear how landscapes of these problems look like and whether they provide a structure that makes local search heuristics reliable and efficient. In a previous research project Barrier tree analysis has been used as a tool that can reveal the local optima structure of these problems. An alternative approach is to use local optima networks that were




**Prerequisites (helpful, not mandatory):** GraphViz, MATLAB, Comninatorial Optimization Algorithms

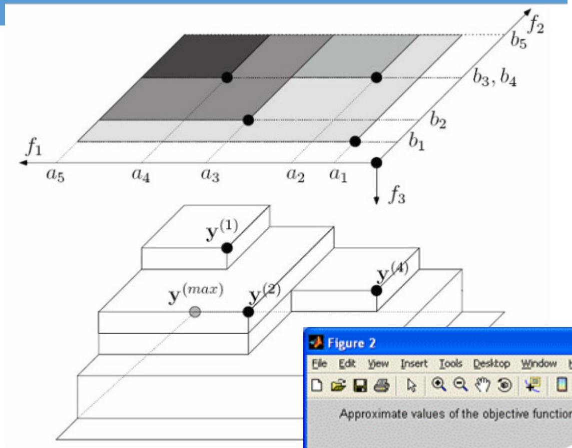
# Research Project: Expected Improvement Algorithm for Global Multiobjective Optimization

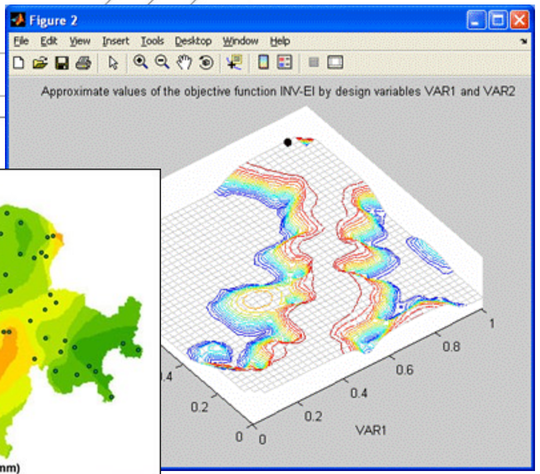
| | |
|---|---|
| Supervisor LIACS: | Prof. dr. Michael Emmerich |
| In cooperation with: | Dr. Andre Deutz |
| Start date: | 2.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |

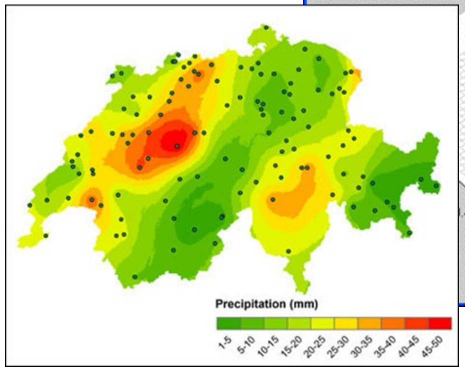
Note: In the project it is possible to Collaborate with researchers in the University of Applied Science, Cologne Gummersbach, Germany.

**Summary: Prerequisites (helpful, not mandatory):** Evolutionary Multiobjective Optimization Algorithms (EMOAs) are recently widely used to approximate sets of efficient solutions in problems with multiple, possibly conflicting, objective functions. The expected improvement (EI) is a measure used in optimizing systems based on time expensive black-box evaluations to estimate the distance of a yet unevaluated point from the optimum. Since short, a precise computation has been made available by researchers of LIACS for this widely used measure. However, there is work to be done to test this algorithm on different multi-objective problems and to assess possibilities to integrate constraints.

- Interest and basic knowledge in statistical distributions, algorithm complexity, and calculus
- MATLAB and C/C++
- Data structures and algorithmics



# Master Project: Multi-point expected improvement

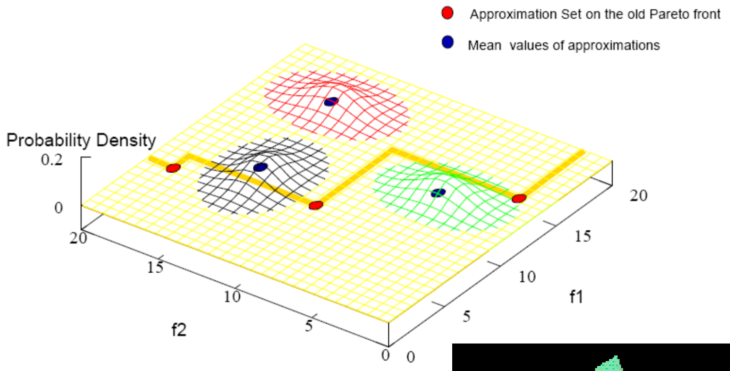
| | |
|---|---|
| Supervisor LIACS: | Prof. dr. Michael Emmerich |
| In cooperation with: | Dr. André Deutz |
| Start date: | 2.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |

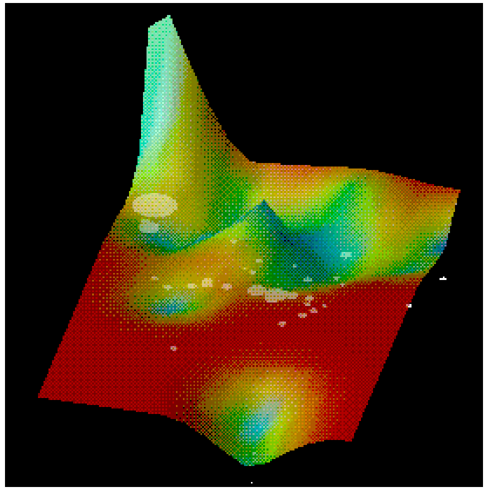
In collaboration with Newcastle University, Dr. Iryna Yevseyeva

**Summary: Expected Hypervolume improvement is a measure that can be used in experimental optimization in order to find the most promising candidate for the next experiment. It is a global optimization method for time consuming optimization. The goal is to find a method that selects multiple points, in case the experiments can be conducted in parallel. The project adresses a ‚hot topic' in global optimization and a novel approach to solve this problem will be implemented, tested and possibly improved.**

**Prerequisites (helpful, not mandatory):**

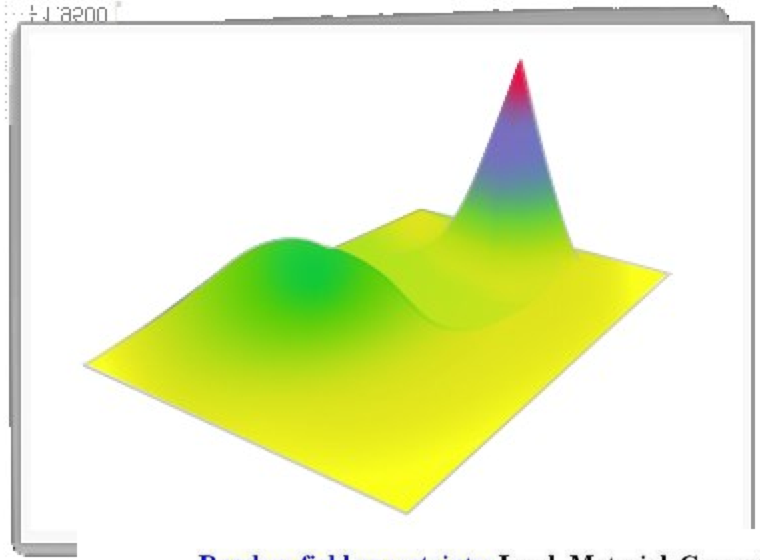
**Interest in optimization and mathematical software, MATLAB**



# Research Project: Dynamic Updates in Evolutionary Robust Optimization

| | |
|---|---|
| Supervisor LIACS: | Prof. dr. Michael Emmerich |
| In cooperation with: | Dr. Jürgen Branke |
| Start date: | 2.9.2014 (immediately) |
| Contact: | emmerich@liacs.nl |



**Summary: Prerequisites (helpful, not mandatory):**

Robust design optimization deals with finding stable optima. Testing the stability of an optimum can be extremely time consuming. Sampes can be recycled and overlap can be exploited in order to reduce complexity. Gibbs samplers can make this process very efficient. Dynamic updates are used in combinatorial optimization but not yet in robust optimization, and a study of this strategy might yield a high impact result.

- Interest and basic knowledge in statistical distributions, algorithm complexity, and calculus
- MATLAB , or C
- evolutionary algorithms