# PhD. Proposal: A Universal Optimiser for Binary Valued Functions

Kevin Swingler

March 28, 2017


## 1 Abstract

Mixed order hyper networks (MOHNs) [7] are neural network models capable of representing any function in $f : \{-1, 1\}^n \rightarrow \mathbb{R}$. They are linear parameter models, which means there are convex cost functions for estimating the weight parameters and they can learn from a noise free data sample equal in size to the number of weights in the network. These qualities make then highly suitable for use as fitness function models when attempting to optimse a function [5]. The proposed thesis will investigate and compare methods for searching a MOHN once it has been learned, making use of the explicit way in which the function expresses the constraints to be optimised.


## 2 Introduction

Heuristic optimisation is concerned with the design and development of heuristic algorithms capable of finding the inputs to a given function that optimise some aspect of its output (maximise or minimise it, for example). Most methods rely on evaluating one or more candidate solutions at a time and allowing the output from those solutions to guide the generation of new solutions. Where the function being searched is expensive to evaluate, it is important to minimise the number of candidate solutions processed. One way to do this is to build a statistical model of the function and optimise that instead. This is known as fitness function modelling. For functions that can be modelled in fewer evaluations than it takes to optimise them, this can provide considerable improvements in the speed with which solutions are found. However, it still leaves the problem of how to search the model. Further gains can be made if the model's structure can be used to guide that search.

Mixed order hyper networks (MOHNs) are universal function models in $f : \{-1, 1\}^n \rightarrow \mathbb{R}$ and can learn a function in a number of noise free samples equal to the number of parameters in the model. This places a bound on the number of function evaluations required to learn the function. A MOHN representation allows the easy extraction of a set of weighted weak constraints that make the form and complexity of the function being optimised explicit. As the MOHN can represent any real valued function of binary inputs, it can also represent any optimisation problem as a set of such constraints. The constraints may then be used to guide the search for an optimal input. Swingler has presented experimental results that show a MOHN can learn and optimise a number of benchmark functions in fewer fitness evaluations than state of the art methods such as BMDA [2] DEUM [4], BOA [1], hBOA [9], and Boltzmann EDAs [3].

# 3 Proposal

Let $X_i, i = 1 \dots n$ be a vector of $n$ binary variables. A MOHN can represent any real valued fitness function, $f(X)$ as a set of $m$ weighted weak constraints in the form

$$\text{sign}(\prod_{i \in \mathbf{I}_j} X_i) = \text{sign}(\omega_j),\ j = 1 \dots m \tag{1}$$

where $\omega_j \in \mathbb{R}, j = 1 \dots m$ represents a set of weighted constraints and $\mathbf{I}_j, j = 1 \dots m$ represents a set of index sets that define which elements of $X$ are involved in each constraint. The value of each $\omega_j$ specifies the importance (or weight) of the constraint. Constraints can conflict, so there are many functions in which only a subset of the constraints can be satisfied. Maximising the output of the function is equivalent to satisfying the set of compatible constraints with the largest sum of weights.

The proposed PhD will investigate and compare methods for satisfying the constraints as defined in equation 1. There is also scope for further investigation of hybrid methods for learning and searching MOHNs with the goal of minimising the number of fitness function evaluations. The work is well suited to PhD study as it is very clearly defined, certainly not trivial, and the existing literature is easy to identify. The use of statistical models in optimisation represents an active research field, though the majority of the work involves estimation of distribution algorithms (EDAs).

The student will compare existing optimisation methods and similar constraint frameworks such as weighted MaxSat. This will lead to a number of different algorithms being proposed, implemented and tested. A large set of benchmark problems from the existing literature will be used to test the algorithms and a small number of real world examples will also be addressed.

# 4 Motivation

The motivation for using fitness functions and EDAs is well documented [8]. They can reduce the number of evaluations needed and transfer the work of searching to a faster statistical model. Further gains can be made if the structure of the model facilitates the search. The bulk of current work in the field involves EDAs, but recent results have shown that MOHNs can solve a number of benchmark problems in fewer (in some cases, orders of magnitude fewer) fitness evaluations that these methods.

Additional motivation for the MOHN approach comes from its universal nature. There is no function in $f : \{-1, 1\}^n \to \mathbb{R}$ that cannot be represented as a set of constraints in the form defined in equation 1. Good methods for solving problems in that form will be widely applicable. Not all functions may be *efficiently* specified in this way, and more work is needed on identifying appropriate functions. The project fits well withing the data science theme of the computing science and maths strategy.

# 5 Supervision

Kevin Swingler has just completed a PhD on the subject of MOHNs. Many of the findings cited above come from that work. His thesis has led to ten publications in journals, conferences and book chapters. Two of the conference papers were short listed for best student paper and one of them won that award [6].

## References


[1] Matthew Alden and Risto Miikkulainen. Marleda: Effective distribution estimation through markov random fields. Technical Report TR-13-18, Department of Computer Science, The University of Texas at Austin, Austin, TX, November 2013.

[2] M. Pelikan and H. Mühlenbein. The bivariate marginal distribution algorithm. In R. Roy, T. Furuhashi, and P. K. Chawdhry, editors, *Advances in Soft Computing - Engineering Design and Manufacturing*, pages 521–535, London, 1999. Springer-Verlag.

[3] Malte Probst and Franz Rothlauf. Deep Boltzmann machines in estimation of distribution algorithms for combinatorial optimization. *arXiv preprint arXiv:1509.06535*, 2015.

[4] Siddhartha Shakya, John McCall, Alexander Brownlee, and Gilbert Owusu. Deum - distribution estimation using markov networks. In Siddhartha Shakya and Roberto Santana, editors, *Markov Networks in Evolutionary Computation*, volume 14 of *Adaptation, Learning, and Optimization*, pages 55–71. Springer Berlin Heidelberg, 2012.

[5] K Swingler. *Mixed Order Hyper Networks for Function Approximation and Optimisation.* PhD thesis, Stirling University, 2017.

[6] Kevin Swingler. A comparison of learning rules for mixed order hyper networks. In *Proceedings of the 7th International Joint Conference on Computational Intelligence (IJCCI 2015) - Volume 3: NCTA, Lisbon, Portugal, November 12-14, 2015.*, pages 17–27, Setubal, 2015. ScITePress.

[7] Kevin Swingler and Leslie S. Smith. Training and making calculations with mixed order hyper-networks. *Neurocomputing*, (141):65–75, 2014.

[8] Yoel Tenne and Chi-Keong Goh. *Computational intelligence in expensive optimization problems*, volume 2. Springer Science & Business Media, 2010.

[9] Gabriele Valentini, Luigi Malagò, and Matteo Matteucci. Optimization by l1-constrained markov fitness modelling. In *Learning and Intelligent Optimization*, pages 250–264. Springer, 2012.