Book of Abstract

1st International Workshop on Interpretability and Explainability in Optimization

Session 1: Multiobjective Optimization

Explainable Interactive Multiobjective Optimization: The What, The How, and Why You Should Care (Giovanni Misitano)

Real-life decision-making often consist of balancing various conflicting criteria and making compromises. Such problems can be modeled as multiobjective optimization problems. Solving these problems involves optimizing simultaneously multiple conflicting objective functions. Instead of just a single optimum, the result of multiobjective optimization is often a set of so-called Pareto optimal solutions. These solutions cannot be fully ordered from best to worst, and represent different trade-offs between the objective functions. It is up to a decision maker, a domain expert, to compare various Pareto optimal solutions and select the best one to be implemented. This best solution depends on the subjective preferences of the decision maker. There are various types of multiobjective optimization methods that incorporate preferences into the optimization process. In this talk, we will focus specifically on interactive multiobjective optimization methods.

Interactive methods allow the decision maker to iteratively provide preference information and see promising solution candidates during an optimization process. The decision maker is able to refine their preferences according to what is available. Thus, interactive methods allow a decision maker to explore and learn about the available solutions and their own preferences. However, interactive multiobjective optimization methods often lack in their capabilities to support decision makers in various aspects related to the interactive optimization process and decision-making. For instance, decision makers may lack support when providing their preferences, or understanding how their given preferences relate to the solutions found by an interactive method. In other words, to a decision maker, interactive methods may seem like opaque-boxes.

To address these challenges, the concept of explainability is adapted from the field of artificial intelligence to the field of multiobjective optimization. This talk will introduce and discuss the emerging field explainable interactive multiobjective optimization, its current status, and what the future holds. By incorporating explainability, interactive multiobjective optimization methods are able to provide better support to decision makers, which in turn can better help them understand the optimization process, provide deliberate preference information, justify the selected solutions, and more. Advancing the novel field of explainable interactive multiobjective optimization can result in innovations and breakthroughs in how we see and approach decision-making in the near future.

An Estimate-and-Optimize Method for Interpretable Inverse Multiobjective Optimization (Nuria Gómez Vargas)

Decision-making in real-world applications often involves multiple competing objectives, with decision-makers applying their own preferences to balance trade-offs. Inverse Multiobjective Optimization (IMO) aims to infer both the underlying objective functions and the implicit preferences that drive observed decisions. In this work, we propose a novel approach that integrates clustering into inverse optimization to group decision-makers based on their preference structures rather than their observed decisions. Our method is an optimization-based clustering IMO framework that minimizes optimality gaps—the difference between the objective value of observed decisions and the optimal value under the inferred preference structure—resulting in Mixed-Integer Quadratic Programs (MIQPs). To enhance computational efficiency, we derive a heuristic that provides a warm-start solution for the global optimization models, improving convergence and solution quality. Additionally, we incorporate interpretability requirements to ensure that inferred preferences are meaningful and align with domain knowledge. We validate our approach with computational experiments on a real-world diet recommendation problem, demonstrating its ability to uncover interpretable and robust decision-making patterns.

Automatic Explanations of Computation Results with Value Decompositions and Dominating Sets (Prashant Kumar)

In an increasingly computation-driven world, algorithms and mathematical models significantly impact decision making across various fields. To foster trust and understanding, it is crucial to provide users with clear and concise explanations of the reasoning behind the results produced by computational tools, especially when recommendations appear counter intuitive. Legal frameworks in some countries have acknowledged the importance of explainability by including the “right to explanation” in their legislation. During my PhD I developed techniques for explaining the results of algorithms, thereby enhancing transparency and increasing the trustworthiness of computational systems.

My research addresses users' contrastive questions, such as "why did the algorithm produce result X instead of Y?", by employing contrastive explanations. These explanations compare the actual result (fact) with a hypothetical alternative (foil), which assists users in better understanding the algorithm's output. The explanations are assembled from a fine-grained representation of the result value, which requires a decomposition of the input values for the algorithm to be explained. For reasons explained in my work, these explanations are called MDS explanations.

I applied the MDS explanation technique to dynamic programming, combinatorial optimization, and multi-criteria decision-making methods (MCDM), adapting the approach to generate explanations specific to each domain. I also introduced robustness as a measure for distinguishing MDS explanations. A so-called Most Robust Explanation (MRE) is least likely to change when altering underlying values. I present an algorithm for computing MRE. Finally, I introduce a domain-specific language (DSL) embedded in Haskell for encoding and solving multi-attribute, multi-layered decision-making problems. Additionally, the DSL also generates explanations detailing why a particular solution is superior to alternatives.

Session 2: Interpretable Algorithms and Decisions

Interpretable Policies for Markov Decision Processes (Daniël Vos)

Markov Decision Processes (MDPs) can be used to model a wide variety of important problems. However, we typically solve them by mapping every state to an action, which is difficult for humans to interpret when the MDP contains many states, or through a complex policy that maps states to actions. In this talk we discuss methods for optimizing policies as decision trees that are restricted in complexity. Such policies allow for easier interpretability and verifiability.

Interpretable Surrogates for Optimization (Sebastian Merten)

An important factor in the practical implementation of optimization models is their acceptance by the intended users. This is influenced by various factors, including the interpretability of the solution process. A recently introduced framework for inherently interpretable optimization models proposes surrogates (e.g., decision trees) of the optimization process. These surrogates represent inherently interpretable rules for mapping problem instances to solutions of the underlying optimization model. In contrast to the use of conventional black-box solution methods, the application of these surrogates thus offers an interpretable solution approach. Building on this work, we investigate how we can generalize this idea to further increase interpretability while concurrently giving more freedom to the decision maker. We introduce surrogates which do not map to a concrete solution, but to a solution set instead, which is characterized by certain features. Furthermore, we address the question of how to generate surrogates that are better protected against perturbations. We use the concept of robust optimization to generate decision trees that perform well even in the worst case. For both approaches, exact methods as well as heuristics are presented and experimental results are shown. In particular, the relationship between interpretability and performance is discussed.

Explainability in Hyper-Heuristics (Edward Keedwell)

Selection hyper-heuristics combine machine learning and optimisation methods to create algorithms that adapt their strategy to new problems by selecting appropriate (sequences of) low-level heuristics to apply and their parameterisations. The information learned in this process provides a rich dataset of the interactions between algorithm and problem domain(s) which, when combined with methods of explainability, can provide an understanding of algorithm and low-level heuristic efficacy and the algorithm-problem nexus. In this talk I will describe some recent research on hyper-heuristics and methods that have been used to explain the decisions made in both online and offline versions of these algorithms.

Session 3: XAI and AI for Explainability in Optimization

Explainability in Credit Risk Modeling using LLMs with Tabular and Network Data (María Óskarsdóttir)

While machine learning models such as Extreme Gradient Boosting (XGBoost) and Graph Neural Networks (GNNs) offer strong predictive performance, their opaque nature raises concerns about transparency and regulatory accountability. In this work, we conduct a comparative study of explainability in credit risk modelling across two modalities: tabular data and network-structured data. Using Freddie Mac loan-level data, we apply SHapley Additive Explanations (SHAP) to interpret XGBoost predictions and GNNExplainer to interpret GAT-based GNN outputs. To improve accessibility of these explanations, we employ Large Language Models (LLMs) to generate textual narratives from SHAP and GNNExplainer outputs. We evaluate three LLMs across nine configurations spanning both explanation sources and model scales. Explanation quality is benchmarked using automatic metrics and also through human evaluation from both general users and credit risk professionals. Our results highlight trade-offs between model scale, explanation source, and user perception, offering practical insights into the deployment of explainable machine learning systems in financial domains.

Collective LIME: Enhancing the Explainability of the Explainer (Dolores Romero Morales)

Local Interpretable Model-Agnostic Explanations (LIME) is a popular tool in the field of Explainable Artificial Intelligence, to shed light on black-box machine learning models. Given a prediction model and an instance, LIME builds a surrogate linear model which yields similar predictions around the instance. When LIME is applied to a group of instances, independent linear models are obtained, often overlooking global properties, such as smoothness and cost-sensitive feature selection.

In this talk we propose a novel framework, called Collective LIME (CLIME), where the surrogate models built for the different instances are linked, being smooth with respect to the coordinates of the instances. With this collective approach, CLIME enables one to control global sparsity, i.e., which features are used ever, even if sparse models are built for each instance. In addition, CLIME builds Generalized Linear Models as surrogates, allowing us to address with the very same methodology different prediction tasks: classification, regression, and regression of counting data. We will end the talk illustrating our approach on a collection of benchmark datasets.

Coherent Local Explanations for Mathematical Optimization (Daan Otto)

The surge of explainable artificial intelligence methods seeks to enhance transparency and explainability in machine learning (ML) models. At the same time, there is a growing demand for explaining decisions taken through complex algorithms used in mathematical optimization. One promising idea to do this is to use a method like LIME, which fits an explainable ML model that locally approximates the behavior of the black-box ML model. This method can be used to analyze components of the optimization model as well (e.g., the objective function value or the decision variables). Although these ML-based methods are effective and model-agnostic, they usually do not take into account the structure of the underlying optimization problem. While the objective value and the corresponding solution are closely intertwined due to the problem's structure, this relation is not taken into account when approximating both components independently by an ML model.

In response to this need, we introduce Coherent Local Explanations for Mathematical Optimization (CLEMO). CLEMO provides explanations for multiple components of optimization models, the objective value and decision variables, which are coherent with the underlying model structure. Our sampling-based procedure can provide explanations for the behavior of exact and heuristic solution algorithms using regression models. The effectiveness of CLEMO is illustrated by experiments for the shortest path problem, the knapsack problem, and the vehicle routing problem. Currently, we are extending CLEMO to provide explanations using decision trees.

Session 4: Applications and Case Studies

Explanations for an Industrial Workforce Allocation Problem (Ignace Bleukx)

In this talk, we will consider an industrial workforce allocation problem from the aircraft manufacturing industry. The setting of this problem is to schedule and allocate a set of logistical tasks to worker-teams in an efficient and fair manner. However, throughout the working shift, unexpected events such as adverse weather conditions or logistical breakdowns may invalidate the precomputed schedule. We investigate how to assist the planning-team with resolving such disruptions in the schedule by computing classical explanation methods such as MUSes and MCSes, and we propose several alternative schedules to automatically deal with the disruption.

Estimating Maintenance Cost of Offshore Substations: A Case Study for Interpretability and Explainability in Optimization (Solène Delannoy-Pavy)

France aims to deploy 45 GW of offshore wind capacity by 2050. Both the ownership and maintenance of the offshore substations that link these farms to the grid lies with the French Transmission System Operator (TSO). In the event of an unscheduled substation shutdown, the TSO must pay significant penalties to producers. Failures when weather conditions prevent access to the station can quickly snowball into huge losses. It is therefore crucial to estimate the expected penalties associated with maintenance strategies. This enables informed design choices and more effective management of operational risks. Unfortunately, given the novelty of these assets, there is limited information regarding the reliability of offshore substations.

We present a decision support tool to estimate maintenance costs associated with various strategic decisions, such as substation design selection or stock management policy definition. Given the potential costs incurred and the variety of professions involved in the decision-making process, it is essential for the model to be both interpretable and explainable. We collaborated with technical experts with the aim of representing operational insights and rules using a simple model.

We model the problem as a Markov Decision Process, where each state reflects both the asset's degradation level and ongoing maintenance, and actions represent maintenance decisions. The objective is to optimize the maintenance schedule to minimize penalties, which are proportional to curtailed energy when capacity is limited. To account for the impact of external conditions, we incorporate weather scenarios which impact both power and the feasibility of maintenance operations. Our approach uses a multihorizon stochastic optimisation framework that combines a bimonthly strategic horizon for forward planning with a daily operational horizon to capture how penalties evolve under uncertain weather conditions.

Impactful Optimization by Constraining Oneself to Transparency, Fairness and Explainability (Frans de Ruiter)

In this talk we present how successful research in operations research can be performed in practice through transparancy, fairness and explainability. We investigate how well-established approaches for operations research might have a hard time realizing impact in industry and highlight this from personal experiences. We also show how the positive effect on the success and impact you can have by constraining one to transparancy in the benefits, fairness in the outcome and explainability in both model and outcome. We present an adjusted approach for operations research in practice emphasizing business-needs and agility. Even though research is not a priority in this, we show how research results and output can follow even if the starting point are basic approaches. The talk incorporates selected examples from the last 10 years building on experience in both academia and industry during consultancy projects and in semiconductor industry.

Session 5: Counterfactual Explanations

Counterfactual Explanations for Unsatisfiable Producer/ Consumer Problems (Helmut Simonis)

Interactive constraint systems often suffer from infeasibility (no solution) due to conflicting user constraints. A common approach to recover feasibility is to eliminate the constraints that cause the conflicts in the system. This approach allows the system to provide an explanation as: "if the user is willing to drop some of their constraints, there exists a solution". However, this form of explanation might not be very informative. A counterfactual explanation is a type of explanation that can provide a basis for the user to recover feasibility by helping them understand what changes can be applied to their existing constraints rather than removing them. We propose an efficient approach NOPROPCOUNTERFACTUALXPLAIN to find counterfactual explanations for infeasible problems. We also propose a version of this algorithm which takes into account preferences called PREFNOPROPCOUNTERFACTUALXPLAIN. We showcase it's usability in a real-world scenario using the producer/consumer constraint which is useful in problems which involve resource allocation.

Counterfactual Explanations for Integer Linear Optimization (Jannis Kurtz)

In recent years, there has been a rising demand for transparent and explainable AI models. Although significant progress has been made in providing explanations for machine learning (ML) models, this topic has not received the same attention in the Operations Research (OR) community. To tackle this issue we introduce the concept of counterfactual explanations and show how it can be used to calculate explanations for linear integer optimization problems. We show that calculating weak and strong CEs is Σ_2^p-hard but can often be solved in reasonable time by problem-specific algorithms.

Relative Explanations for Contextual Problems with Endogenous Uncertainty: An Application to Competitive Facility Location (Jasone Ramírez-Ayerbe)

In this talk, we consider contextual stochastic optimization problems subject to endogenous uncertainty, where the decisions affect the underlying distributions. To implement such decisions in practice, it is crucial to ensure that their outcomes are interpretable and trustworthy. To this end, we compute relative counterfactual explanations, providing practitioners with concrete changes in the contextual covariates required for a solution to satisfy specific constraints. Whereas relative explanations have been introduced in prior literature, to the best of our knowledge, this is the first work focused on problems with binary decision variables and subject to endogenous uncertainty. We propose a methodology that uses Wasserstein distance as regularization and to compute a lower bound. It leads to a drastic reduction in computation times, compared to the unregularized counterpart. We illustrate the method using a choice-based competitive facility location problem, and present numerical experiments that demonstrate its ability to efficiently compute sparse and interpretable explanations.

Session 6: From Data to Decisions

On a Computationally Ill-Behaved Bilevel Problem with a Continuous and Nonconvex Lower Level (Johannes Thürauf)

It is well known that bilevel optimization problems are hard to solve both in theory and practice. We highlight a further computational difficulty when it comes to solving bilevel problems with continuous but nonconvex lower levels. Even if the lower-level problem is solved to epsilon-feasibility regarding its nonlinear constraints for an arbitrarily small but positive epsilon, the obtained bilevel solution as well as its objective value may be arbitrarily far away from the actual bilevel solution and its actual objective value. This result even holds for bilevel problems for which the nonconvex lower level is uniquely solvable and its convex constraint set satisfies Slater's constraint qualification for all feasible upper-level decisions. We further illustrate that the nonlinearities in the lower level are the key reason for the observed bad behavior by showing that linear bilevel problems behave much better at least on the level of feasible solutions. Thus, our result shows that computational bilevel optimization with continuous and nonconvex lower levels and the interpretability of the corresponding results needs to be done with great care.

Data-driven Explainable Mathematical Optimization Including Feature Selection (Kevin Aigner)

Mathematical optimization is a powerful tool for solving complex real-world problems, but its acceptance is often hindered by a lack of trust and the perception of solutions as black boxes. To address this challenge, we introduce explainability as an additional evaluation criterion alongside solution quality. Our approach justifies optimized solutions by relating them to similar solutions from past problem instances, thereby enhancing transparency and trust. We formulate this explainable framework within mathematical programming, analyze its computational complexity, and identify tractable special cases such as the explainable shortest-path problem. Moreover, we develop a feature-selection methodology to establish meaningful similarity measures between instances, using mixed-integer programming. Numerical experiments on artificial and real-world datasets demonstrate that explainability can be achieved at low cost while improving solution acceptance in practice.

Via Classical Nonlinear Optimization to Machine Learning and Back (Krzysztof Postek)

Before optimizing any market participants' decisions, one must first model complex other players' dynamics in response to these decisions. This requires (i) data and (ii) economically/business meaningful assumptions on demand and reaction curves. Often, economic coefficients must also satisfy constraints, e.g., hierarchical relations in customer price elasticities. This combination of large data and numerous constraints creates an estimation challenge: the scale calls for ML-style computational power, but the constraints exclude most ML algorithms, making classical optimization indispensable. In this talk, we highlight this striking research gap in joint elasticity estimation and sketch how it can be addressed in a fast-paced development environment, pointing that this difficult problem is only a prelude to the much more complex decision problem.