Abstracts of talks for 1st ACDA Workshop at Aussois
(in the order of submission)

Julian Shun

Parallel batch-dynamic algorithm for approximate k-core decomposition

In this talk, we present the first parallel batch-dynamic algorithm for approximate k-core decomposition that is efficient in both theory and practice. Our algorithm is based on our novel parallel level data structure, inspired by the sequential level data structures of Bhattacharya et al. and Henzinger et al. Given a graph with n vertices and a batch of B updates, our algorithm maintains a (2 + epsilon)-approximation of the coreness values of all vertices (for any constant epsilon > 0) in O(B log^2(n)) amortized work and O(log^2(n) loglog(n)) span (parallel time) with high probability. We implement and experimentally evaluate our algorithm, and demonstrate significant speedups over state-of-the-art serial and parallel implementations for dynamic k-core decomposition. Using our parallel level data structure, we also obtain new parallel batch-dynamic algorithms for low out-degree orientation, maximal matching, clique counting, and graph coloring.

David Shmoys

Algorithmic Tools for Congressional Districting: Fairness via Analytics

The American winner-take-all congressional district system empowers politicians to engineer electoral outcomes by manipulating district boundaries. To date, computational solutions mostly focus on drawing unbiased maps by ignoring political and demographic input, and instead simply optimize for compactness and other related metrics. However, we maintain that this is a flawed approach because compactness and fairness are orthogonal qualities; to achieve a meaningful notion of fairness, one needs to model political and demographic considerations, using historical data.

We will discuss two papers that explore and develop this perspective. In the first (joint with Wes Gurnee), we present a scalable approach to explicitly optimize for arbitrary piecewise-linear definitions of fairness; this employs a stochastic hierarchical decomposition approach to produce an exponential number of distinct district plans that can be optimized via a standard set partitioning integer programming formulation. This enables the largest-ever ensemble study of congressional districts, providing insights into the range of possible expected outcomes and the implications of this range on potential definitions of fairness. In the second paper (joint with Nikhil Garg, Wes Gurnee, and David Rothschild), we study the design of multi-member districts (MMDs) in which each district elects multiple representatives, potentially through a non-winner-takes-all voting rule (as currently proposed in H.R. 4000). We carry out large-scale analyses for the U.S. House of Representatives under MMDs with different social choice functions, under algorithmically generated maps optimized for either partisan benefit or proportionality. We find that with three-member districts using Single Transferable Vote, fairness-minded independent commissions can achieve proportional outcomes in every state (up to rounding), and this would significantly curtail the power of advantage-seeking partisans to gerrymander.

Jessica Shi

Theoretically and Practically Efficient Parallel Nucleus Decomposition

This work studies the nucleus decomposition problem, which has been shown to be useful in finding dense substructures in graphs. We present a novel parallel algorithm that is efficient both in theory and in practice. Our algorithm achieves a work complexity matching the best sequential algorithm while also having low depth (parallel running time), which significantly improves upon the only existing parallel nucleus decomposition algorithm (Sariyuce et al., PVLDB 2018). The key to the theoretical efficiency of our algorithm is the use of a theoretically-efficient parallel algorithms for clique listing and bucketing. We introduce several new practical optimizations, including a new multi-level hash table structure to store information on cliques space-efficiently and a technique for traversing this structure cache-efficiently. On a 30-core machine with two-way hyper-threading on real-world graphs, we achieve up to a 55x speedup over the state-of-the-art parallel nucleus decomposition algorithm by Sariyuce et al., and up to a 40x self-relative parallel speedup. We are able to efficiently compute larger nucleus decompositions than prior work on several million-scale graphs for the first time.

Michael Bender

Online Parallel Paging and Green Paging

We give provably good algorithms for online paging in multicore systems. We solve this decades-old open problem by showing a connection to the problem of green paging.

Johannes Langguth

ML Accelerator Hardware: A Model for Parallel Sparse Computations ?

Recently, dedicated accelerator hardware for machine learning applications such as the Graphcore IPUs and Cerebras WSE have evolved from the experimental state into market-ready products, and they have the potential to constitute the next major architectural shift after GPUs saw widespread adoption a decade ago.

In this talk we will present the new hardware along with implementations of basic graph and matrix algorithms and show some early results on the attainable performance, as well as the difficulties of establishing fair comparisons to other architectures. We follow up by discussing the wider implications of the architecture for algorithm design and programming , along with the wider implications of adopting such hardware.

Sivan Toledo

Combinatorial Problems and Algorithms in Robust Estimation

The talk will discuss a few combinatorial problems and algorithms in robust estimation. Estimation is the process of finding the parameters of some system or model from noisy observations of the system. Usually, the parameters cannot be observed directly. When the system is linear and the noise is Gaussian, linear least squares gives the optimal solution (in some specific technical senses). Other assumptions on the distribution of noise lead to linear programming problems. In *robust* estimation, the noise is not constrained to a known statistical distribution. For example, we may assume that most of the observations are only contaminated by well-behaved Gaussian noise, but some by arbitrary errors. The talk will survey some hardness results regarding such problems and some heuristic algorithms, some very old and very useful in practice, and some more recent, based on my own recent work.

Henning Meyerhenke

Faster Greedy Optimization of Resistance-based Graph Robustness

The total effective resistance, also called the Kirchhoff index, provides a robustness measure for a graph G. We consider the optimization problem of adding k new edges to G such that the resulting graph has minimal total effective resistance (i. e., is most robust). The total effective resistance and effective resistances between nodes can be computed using the pseudoinverse of the graph Laplacian. The pseudoinverse may be computed explicitly via pseudoinversion; yet, this takes cubic time in practice and quadratic space. We instead exploit combinatorial and algebraic connections to speed up gain computations in established generic greedy heuristics. Moreover, we leverage existing randomized techniques to boost the performance of our approaches by introducing a sub-sampling step. Our different graph- and matrix-based approaches are indeed significantly faster than the state-of-the-art greedy algorithm, while their quality remains reasonably high and is often quite close. Our experiments show that we can now process large graphs for which the application of the state-of-the-art greedy approach was infeasible before. As far as we know, we are the first to be able to process graphs with 100K+ nodes in the order of minutes.

Joint work with Maria Predari (HU Berlin) and Rob Kooij (TU Delft).

Lutz Oettershagen

Inferring Tie Strength in Temporal Networks

Inferring tie strengths in social networks is an essential task in social network analysis. Common approaches classify the ties as weak and strong ties based on the strong triadic closure (STC). The STC states that if for three nodes, A, B, and C, there are strong ties between A and B, as well as A and C, there has to be a (weak or strong) tie between B and C. So far, most works discuss the STC in static networks. However, modern large-scale social networks are usually highly dynamic, providing user contacts and communications as streams of edge updates. Temporal networks capture these dynamics. To apply the STC to temporal networks, we first generalize the STC and introduce a weighted version such that empirical a priori knowledge given in the form of edge weights is respected by the STC. The weighted STC is hard to compute, and our main contribution is an efficient 2-approximative streaming algorithm for the weighted STC in temporal networks. As a technical contribution, we introduce a fully dynamic 2-approximation for the minimum weight vertex cover problem, which is a crucial component of our streaming algorithm. Our evaluation shows that the weighted STC leads to solutions that capture the a priori knowledge given by the edge weights better than the non-weighted STC. Moreover, we show that our streaming algorithm efficiently approximates the weighted STC in large-scale social networks.

Sebastian Christodoulou

Smoothing Discontinuous Concatenated Functions

Piecewise continuous functions can be regarded as a sets of continuous subfunctions defined on parts of the domain with a discontinuous, abrupt transition between them. Smoothing such subfunctions requires knowledge about the location of these discontinuities. When dealing with a concatenated function, locating the discontinuities or the subfunction associated to any location, generally requires function inverses, and thus becomes unrealistic for generic code. The introduced approach instead makes all smoothing-related decisions based on the local discontinuity of the components of a concatenation. This

results in a globally smoothed version of the discontinuous concatenation. The complexity of evaluating a concatenation of n locally smoothed functions is exponential in the number n of concatenation components. The number of evaluations in the average case can be significantly reduced by using a heuristic, which only evaluates relevant subfunctions that contribute to the result at any input. The complexity of the introduced method does not depend on the input-dimension of the function. Eventually, methods to obtain smooth derivatives for discontinuous concatenations are discussed. For the case of programming-code, smooth derivatives are obtained by combining our smoothing-method with Automatic Differentiation.

Michele Scquizzato

Matching on the line admits no $o(\sqrt{\log n})$-competitive algorithm

We present a simple proof that the competitive ratio of any randomized online matching algorithm for the line is $\Omega(\sqrt{\log n})$. This is the first superconstant lower bound for the problem, providing a negative answer to a question open for the last quarter-century.

Nate Veldt

Greedily Growing a Maximal Independent Set to Approximate Vertex Cover

This talk will cover a simple new randomized 2-approximation algorithm for the minimum weighted vertex cover problem. There are already many fast 2-approximations for vertex cover that are based on iterating through the edges of a graph in arbitrary order, but this approach differs in that it iterates through nodes in a random order and greedily forms a maximal independent set. Although not faster when applied directly to an instance of vertex cover, an implicit implementation of this node-visiting approach leads (somewhat counterintutively) to faster approximation algorithms for certain edge-deletion problems that can be reduced to vertex cover in an approximation preserving way. I'll also discuss a few new connections between vertex cover and a signed graph partitioning objective called correlation clustering, which led to the development of the new vertex cover algorithm.

James Murphy

Density-Driven Path Metrics: Graphs, Manifolds, and Data

Fermat distances (FD) are a class of data-driven metrics based on geodesic paths that penalize passing through low density regions. We demonstrate how certain Fermat distances may be leveraged to design data clustering algorithms that enjoy performance guarantees and robustness to key parameters. We provide provably fast algorithms for FD based on Euclidean nearest neighbor graphs, allowing our approaches to scale quasilinearly in the number of data points. When FD are defined on an i.i.d. sample from a measure supported on a Riemannian manifold M, we prove convergence in the large sample limit of the discrete FD to a particular geodesic defined on M. The interplay between (1) graphs generated from samples and (2) manifolds on which the sampling measure is supported is emphasized throughout, and several lines of ongoing work and open questions will be discussed.

Erik Boman

Graph Partitioning and Randomized Linear Algebra.

We first present Sphynx, a parallel spectral graph partitioner. The computationally most expensive part of spectral partitioning is to compute eigenpairs. We show how randomized linear algebra can quickly give us good approximate eigenpairs. This is sufficient for partitioning highly irregular graphs, such as social networks and web graphs. Joint work with Jennifer Loe and Heliezer Espinoza.

Edgar Solomonik

Efficient Algorithms via Inexact Linear Solvers and Randomized Sampling

We survey a few recent algorithmic advances in numerical and combinatorial problems, based on new analysis of randomized approximation algorithms. In particular, we propose more efficient sketching schemes for tensor-structured inputs, yielding fast inexact optimization algorithms for tensor decomposition and tensor networks (joint work with Linjian Ma, arXiv:2104.01101/NeurIPS'21 and arXiv:2205.13163). Then, we propose a new inexact solver for sequences of linear systems arising in interior point methods for constrained quadratic programming (joint work with Samah Karim, arXiv:2104.12916/SIMAX'22). Finally, we describe a new, tight analysis of randomized Histogram-based approximate partitioning algorithms for parallel sorting (joint work with Wentao Yang and Vipul Harsh, arXiv:2204.04599).

Florian Kurpicz

Packed and Compressed Hash Tables

PaCHash is a hash table that stores its objects contiguously in an array without intervening space, even if the objects have variable size. In particular, each object can be compressed using standard compression techniques. A small search data structure allows locating the objects in constant expected time. PaCHash is most naturally described as a static external hash table where it needs a constant number of bits of internal memory per block of external memory.

In this talk, I will present PaCHash and its experimental evaluation showing that PaCHash has lower space consumption than all previous approaches even when considering only objects of identical size. I will conclude the talk with open questions regarding the internal data structures' bit vectors structure that we see in practice, which might be exploited to further compress the bit vectors.

Hanna Komlos

Online List Labeling: Breaking the log^2(n) Barrier

The list labeling problem is a classical combinatorial problem with many algorithmic applications. There has long existed a gap between the lower and upper bounds in the most algorithmically interesting part of the problem's parameter space. We present our recent results, which narrow this gap for the first time in nearly 4 decades.

Helen Xu

Optimizing Dynamic Graph Processing with the Locality-First Strategy

This talk will demonstrate how to create parallel data structures for dynamic graph processing that are both theoretically and practically efficient with the "locality-first strategy." That is, they should first understand and exploit locality as much as possible before introducing parallelism.

Zipei Nie

Matrix anti-concentration inequalities with applications

We present an algorithm to solve the sparse linear system Ax=b in R^n in time O(n^2.2716). A key ingredient in our analysis is a polynomial lower bound on minimum singular value of an m * m random matrix M with jointly Gaussian entries, under a polynomial bound on the matrix norm and a global small-ball probability bound. Then we establish two matrix anti-concentration inequalities, which lower bound the minimum singular values of the sum of independent positive semidefinite self-adjoint matrices and the linear combination of independent random matrices with independent Gaussian coefficients. Both are under a global small-ball probability assumption. Finally, we build a lower bound on the minimum singular value for the Krylov space matrix, which leads to the better complexity.

Cédric Chevalier and Sébastien Morais

Coupe: a mesh partitioner

We rely on partitioners such as (Par)Metis, Scotch, or Zoltan for load balancing numerical simulations. As our simulations use meshes, we have begun to write a partitioner framework, Coupe, in the Rust language. This partitioner focuses only on mesh-related algorithms. It implements or provides access to methods to partition numbers, geometry, or simple topologies. At the same time, the idea behind using Rust is to enable efficient and "fearless" shared memory parallelism.

Sherry Li

Combinatorial problems in sparse matrix and sparse tensor applications

In this talk, we will describe the combinatorial algorithms needs in sparse matrix and tensor computations. We will focus on the open problems in the graph preprocessing stages, such as ordering, symbolic factorization, and communication schedule, and particularly the speculations on the multi-GPU design.

Prashant Pandey

Vector quotient filter

Today’s filters, such as quotient, cuckoo, and Morton, have a trade-off between space and speed; even when moderately full (e.g., 50%-75% full), their performance degrades nontrivially. The result is that today’s systems designers are forced to choose between speed and space usage. In this talk, we present the vector quotient filter (VQF). Locally, the VQF is based on Robin Hood hashing, like the quotient filter, but uses power-of-two-choices hashing to reduce the variance of runs, and thus offers consistent, high throughput across load factors. Power-of-two-choices hashing also makes it more amenable to concurrent updates, compared to the cuckoo filter and variants. Finally, the vector quotient filter is designed to exploit SIMD instructions so that all operations have $O(1)$ cost, independent of the size of the filter or its load factor. We show that the vector quotient filter is 2× faster for inserts compared to the

Morton filter (a cuckoo filter variant and state-of-the-art for inserts) and has similar lookup and deletion performance as the cuckoo filter (which is fastest for queries and deletes), despite having a simpler design and implementation. The vector quotient filter has minimal performance decline at high load factors, a problem that has plagued modern filters, including quotient, cuckoo, and Morton. Furthermore, we give a thread-safe version of the vector quotient filter and show that insertion throughput scales 3× with four threads compared to a single thread.

Lionel Eyraud-Dubois

Data Distribution for Symmetric Linear Algebra Kernels

We consider the Cholesky factorization on a parallel distributed architecture. We present several results on the effect of allocation schemes on the volume of communication, showing how to take advantage of the symmetry of the access pattern. We first show that the standard 2D block-cyclic allocation is not adapted to this symmetric operation, and propose a Symmetric Block Cyclic (SBC) allocation scheme which significantly improves both communication volume and performance. We then present a theoretical analysis which provides an improved lower bound on the required communication volume. This analysis leads to the design of another allocation scheme, Triangular Block Cyclic, which further improves over the communication volume of SBC and matches the lower bound. Experimental results within the Chameleon library show improved performance compared to standard distribution schemes.

Cynthia Phillips

Write-optimized algorithms for cybersecurity stream monitoring

We describe data structures/data-management algorithms for monitoring high-speed cyber streams. We wish to find instances of specific patterns where elements arrive slowly over time, hidden among high-speed streams of similar traffic. To answer such standing queries, we must store as much data as possible, to avoid losing partial patterns before the final piece arrives. We must recognize a completed pattern and report it as soon as possible while keeping up with the fast stream of arrivals. To store more stream history, we must carefully manage the movement of data between main (fast) storage and secondary (slower) storage. We describe the problem, how it differs from classic streaming theoretical problems, and give some useful algorithms for finite streams. However cyber streams are effectively infinite, since the system must always be prepared to receive the next item. We describe some of the challenges when moving to managing standing queries on effectively infinite streams.

Oded Schwartz

Towards Accelerating AI using Fast and Feasible Matrix Multiplication

Training deep neural networks increasingly requires large resources. It involves significant time spent on matrix multiplication, typically between 45%-95%. Most current math libraries (for CPU and GPU) and all state-of-the-art hardware accelerators (such as Google's TPU and Intel's / Habana Lab's Gaudi) are based on the cubic-time classic matrix multiplication algorithm, despite more than five decades of research on sub-cubic time algorithms. Why is that? Many of the sub-cubic time algorithms are impractical, as they have large hidden constants in the arithmetic complexity, and enormous minimal applicable size. Yet, recent years have seen encouraging studies addressing these obstacles. In this talk I will review several of them: — We provide a high performance general matrix-matrix multiplication that combines fast base change method and pebbling game based optimization scheme applied to Strassen's algorithm. We reduce arithmetic and communication costs, as well as memory footprint. Our algorithm outperforms DGEMM of Intel's MKL on feasible matrix dimensions starting at n = 1024 and obtains up to nearly ×2 speedup for larger matrix dimensions. — Pan's four decades old fast matrix multiplication

algorithms (based on trilinear aggregation method) have, to date, the lowest asymptotic complexity of all algorithms applicable to matrices of feasible dimensions. However, the large coefficients in the arithmetic cost of these algorithms make them impractical. We reduce these coefficients by 90% − 98%, in some cases down to 2, the same leading coefficient as the classical cubic time algorithm. We show that our results are optimal or close to optimal. — Fast recursive matrix multiplication algorithms call the cubic time classical algorithm on small sub-blocks, as the classical algorithm requires fewer operations on small blocks. We obtain a new algorithm that may outperform the classical one, even on small blocks, by trading multiplications for additions. This algorithm goes against the common belief that the classical algorithm is the fastest for small blocks. Specifically, we obtain an algorithm for multiplying 2 × 2 blocks using only four multiplications. This algorithm seemingly contradicts the lower bound of Winograd (1971) on multiplying 2 × 2 matrices. We provide a new lower bound matching our algorithm for 2 × 2 block multiplication, thus showing that our technique is optimal. Joint work with Yoav Gross, Tor Hadas, and Noa Vaknin

---

Albert-Jan Yzelman

Recent advances and challenges in algebraic programming and hypergraph applications

We first recall the concept of a ``humble programmer'', and contrast it to that of a ``hero programmer''. Classically, the former focuses on achieving high productivity, while the latter focuses on extracting peak performance on a given system. Given that both the complexity of programming novel architectures increases, and that the trend of producing heterogeneous systems that combine multiple such complex architectures is becoming normal, humble programming that achieves scalable performance close to peak on a variety of architectures is becoming a necessity-- yet remains a challenge.

This talk introduces the free and open-source Algebraic Programming (ALP) paradigm as such a candidate solution, and briefly reviews ALP/GraphBLAS specifically. Here, programmers must annotate their code with algebraic information, which the compiler then exploits in optimization as well as the detection of programmer errors and unscalable behaviour. We briefly present our novel nonblocking backend. To cope with humble programmers who ideally would not like to use explicit algebraic annotations, illustrate how ALP generalizes some other well-known models such as vertex-centric programming. Exposing multiple humble programming paradigms via a single humble software stack would resolve the main challenge posed.

In another direction of extending the ALP paradigm, we sketch how classical dense linear algebraic concepts may be incorporated into the ALP paradigm, broadening its usability. We touch on why ALP/Dense requires deeper interactions between compilers and applications, as well as MLIR as a vehicle of capturing and using algebraic knowledge within a compiler stack. We highlight furthermore the attractiveness of such a solution in terms of architecture portability. Finally, we introduce a backend that allows the extraction of HyperDAGs from ALP programs, which will be released with an initial database of example HyperDAGs. After briefly reviewing the notion of a HyperDAG, we share some recent results about HyperDAG partitioning applied to multicore and NUMA-aware scheduling.

Olivier Beaumont

Memory optimization for training in DNNs

In this talk, we will consider some optimization problems related to training in deep neural networks (DNNs). We will show how backpropagation introduces original problems compared to task graphs from linear algebra, in particular in terms of memory management. Due to the limited memory on GPUs, these problems require original trade-offs between memory and computational speed (re-materialization),

memory and use of communication resources (activation and weight offloading). Joint work with Lionel Eyraud Dubois, Alena Shilova, Xunyi Zhao

---

Alex Conway

SplinterDB + Maplets: Improving the Trade-offs in Key-Value Store Compaction Policy

A critical aspect of modern key-value stores is the interaction between compaction policy and filters. Aggressive compaction reduces the on-disk footprint of a key-value store and can improve query performance, but can reduce insertion throughput because it is I/O and CPU expensive. Filters can mitigate the query costs of lazy compaction, but only if they fit in RAM, limiting the scalability of queries with lazy compaction.

In this work, we present a new version of SplinterDB, a state-of-the-art lazily compacted system, which uses maplets, a new type of filter. Maplets enable SplinterDB to decouple the compaction schedules of filters and data, so that filters can be compacted aggressively and data lazily, so that SplinterDB + maplets achieves excellent insertion performance, query performance, space efficiency, and scalability. Maplets can accelerate queries even when they don't fit in RAM, improving scalability to huge datasets.

In our benchmarks, the SplinterDB with maplets matches the insertion performance of unmodified SplinterDB and beats RocksDB by up to 9×. On queries, the SplinterDB with maplets outperforms unmodified SplinterDB and RocksDB by up to 89% and 83%, respectively, and scales gracefully to huge datasets.

Aydın Buluç

Sparse matrices in biology and machine learning

I highlight some of the emerging use cases of sparse matrices outside the domain of numerical solvers. These include computational biology and graph-based machine learning. In particular, I will show that genome assembly for long reads, as well as protein-similarity search and protein family discovery can be done efficiently on distributed-memory supercomputers thanks to mature parallel sparse matrix primitives. Further, sparse matrices power graph neural networks such as graph convolutional networks and graph attention networks.

Kathrin Hanauer

Disjoint Weighted Matchings for Reconfigurable Optical Datacenter Technologies

Reconfigurable optical topologies promise to improve the performance in datacenters by dynamically optimizing the physical network in a demand-aware manner. State-of-the-art optical technologies allow to establish and update direct connectivity between top-of-rack switches within microseconds or less. However, to fully exploit temporal structure in the demand, such fine-grained reconfigurations also require fast algorithms for optimizing the direct connections to establish.

We demonstrate that the problem of offloading a maximum amount of demand to the reconfigurable network can be cast as an edge-disjoint weighted matchings problem, show its NP-hardness, and initiate the study of fast practical algorithms. Altogether, we present and analyze six different algorithms and show in an extensive evaluation on 88 real-world and synthetic traces that we can obtain solutions of extremely high quality very quickly. However, the best choice of algorithm in practice depends not only on

what tradeoff between solution quality and running time is deemed acceptable, but also on the number of optical switches.

Seth Gilbert

To Catch a (Distributed) Thief

Over the last several years we have seen a boom in the development of new Byzantine agreement protocols, in large part driven by the excitement over blockchains and cryptocurrencies. Unfortunately, Byzantine agreement protocols have some inherent limitations: (a) they are very expensive, particularly in terms of bandwidth; (b) correctness tends to depend on strong assumptions, such as unreliable network behavior; and (c) it is impossible to ensure correct operation when more than 1/3 of the processing power in the system is controlled by a single malicious party. These problems are fundamentally intertwined with the problem of Byzantine fault tolerance.

What if, instead of preventing bad behavior by a malicious attacker, we guarantee "accountability," i.e., we can provide irrefutable evidence of the bad behavior and the identity of the perpetrator of those illegal actions? Much in the way we prevent crime in the real world, we can prevent bad behavior in a distributed system: either the protocol succeeds, or alternatively we record sufficient information to catch the criminal and take remedial actions. (Accountability has been increasingly discussed as a desirable property in blockchains like Ethereum to slash stake of cheating users.) The problem is to avoid suspecting correct peers while provably identifying cheating ones.

In this talk, we give an overview of accountability in distributed systems, with a focus on decision problems like consensus. We begin with PeerReview, the first (provably) accountable consensus protocols, and then describe the ABC transformation for making every consensus protocol accountable. Finally, we talk about the underlying theory of accountability and the necessary and sufficient conditions for achieving accountability.

This talk covers joint work by Pierre Civit, Vincent Gramoli, Rachid Guerraoui, Jovan Komatovic, Zarko Milosevic, and Adi Serendinschi, and appeared at ICDCS 2021, IPDPS 2022 (Best Paper), and ICDCS 2022 (Best Paper).

Vanesa Guerrero and Claudia D'Ambrosi

On spline surrogate models and reformulation techniques for MINLPs with separable non-convexities

The vast amount of data available nowadays makes possible to model complex phenomena in accurate ways. However, being able to use these models within a mathematical optimization framework to solve instances of real problems is often a challenge. Many of these data-driven models are `black-box', in the sense that they do not have an explicit mathematical formula which describes it. In other cases, even if an explicit expression exists, including it into a mathematical optimization model may make solving the problem computationally intractable.

In this work we propose the use of a special kind of surrogate models, regression splines, together with some reformulation techniques to deal with `black-box' or too complex functions involved in Mixed Integer Nonlinear Programming (MINLP) problems. The choice of spline functions is not arbitrary. On one hand, they offer a good compromise between accuracy in capturing the main trends in the data and complexity, since they are piecewise polynomials. On the other hand, their functional form allows us to approximate general non-convex MINLPs by a more tractable subclass of problems which can be efficiently solved by customized algorithms.