Optimization Methods for Big Data and Machine Learning

big data, machine learning, numerical optimization, distributed optimization

The increasing generation of vast amounts of data everyday from various sources, including, e.g., internet, social networks, astronomy observations, weather/climate and geo-spatial data, to name a few, creates many new opportunities for extraction of hidden knowledge, but it also poses several new challenges to data analytics in today's Big Data era. Efficient tools, frameworks, and methods for acquisition, modeling/representation, preparation, analysis, and visualization of Big Data remain important challenges. Among those, the design of new mathematical methods applicable for analysis of very large data sets is one of the pressing research challenges. Many data analysis problems, arising, e.g., in machine learning, can be formulated as nonlinear programming (optimization) problems defined over the available data sets. In this context, a key challenge is the development of new optimization algorithms which are able to process huge amounts of data in an admissible time, and which can scale well with the increase of the available computing power and communication bandwidth. Within our group, continuous optimization, in particular distributed and stochastic optimization are main research topics. Specifically, we developed parallel and distributed optimization algorithms parallel which can harness the inherent parallelism over a cluster/network of computing resources and which can adapt to various computing and communication infrastructures. The large volume of data which may not be feasible to manipulate in memory 'as a whole' further motivates stochastic and subsampled methods in this context. The problems we study range from numerical optimization methods for large scale problems, distributed first and second order methods on different types of computer networks, and stochastic optimization, to various applications like machine learning and prediction models in economics, finance and science. We carry out research on the development of state of the art numerical methods, their convergence analysis, convergence and convergence rate guarantees, as well as their efficient implementation in relevant software and frameworks for parallel and distributed processing. We are working on a number of projects which involve multidisciplinary teams and aim at providing practical solutions for several Big Data challenges in many disciplines, including finance, telecom, manufacturing, and bio-systems.

SELECTED PROJECTS

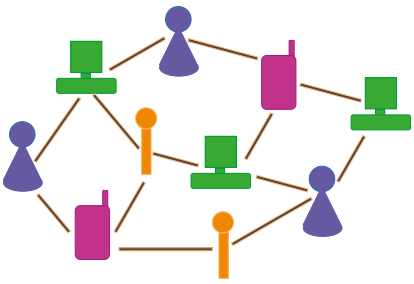
Title: *Industrial-Driven Big Data as a Self-Service Solution - I-BiDaaS*
Type: H2020-ICT-2017-1, Research and Innovative Action
Duration: 2017-2020
Contact person: Dušan Jakovetić

Title: *Big Data Challenges for Mathematics – BIGMATH*
Type: H2020 Marie Sklodowska Curie action, European Industrial Doctorate
Duration: 2018-2022
Contact person: Nataša Krejić

Title: *C4IoT*
Type: H2020-SU-ICT-2018-2020 Innovation Action
Duration: 2019-2021
Contact person: Dušan Jakovetić

COLLABORATIONS

- Tax Administration of the Republic of Serbia: Development of risk factors and machine learning algorithms for detecting tax evasion
- European Consortium for Mathematics in Industry Special Interest Group: Mathematics for Big Data
- Sioux Lime, Eindhoven: Deep learning methods



CONTACT PERSONS

Prof. Dr Nataša Krejić,
natasak@uns.ac.rs
Dr Dušan Jakovetić,
dusan.jakovetic@dmi.uns.ac.rs