# 2023 NCTS Optimization Workshop

**Department of Mathematics**
**National Taiwan Normal University**

**May 17, 2023**

**Sponsored by**

**National Center for Theoretical Sciences, Mathematics Division**
**College of Science, National Taiwan Normal University**
**Department of Mathematics, National Taiwan Normal University**

**Organized by**

**Jein-Shan Chen**

**Table 1: Schedule on May 17, 2023. Place: M212, Mathematics Building**

| | Speaker | Title |
|---|---|---|
| 13:20 – 14:00 | Adil Bagirov | Nonsmooth DC Optimization: Methods and Applications |
| 14:00 – 14:40 | Yeonjong Shin | Active Neuron Least Squares: A Training Method for Multivariate Rectified Neural Networks |
| 14:40 – 15:00 | | *Tea Break* |
| 15:00 – 15:40 | Jan Harold Alcantara | Smoothing Approach for Hyperparameter Learning Using Bilevel Programming |
| 15:40 – 16:20 | Pierre-Louis Poirion | Random Subspace Newton Method for Unconstrained Non-convex Optimization |
| 16:20 – 17:00 | Yoon Mo Jung | Optimization Models for Trend Filtering |
| 17:00 – 17:30 | | *Free Discussion* |

# Smoothing Approach for Hyperparameter Learning Using Bilevel Programming

Jan Harold Alcantara
Continuous Optimization Team
Center for Advanced Intelligence Project
RIKEN, Japan
E-mail: jan.alcantara@dlsu.edu.ph

**Abstract**. In this talk, we focus on the hyperparameter selection of regularizers related to the $\ell_p$ function with $0 < p \leq 1$. We apply a bilevel programming strategy, wherein we need to solve a bilevel problem whose lower-level problem is nonsmooth, possibly nonconvex and non-Lipschitz. We propose bilevel KKT conditions, which are new necessary optimality conditions for the original bilevel program when $p = 1$, and for a relaxed one-level problem when $p < 1$. Moreover, we propose a unified smoothing approach using smoothing functions that belong to the Chen-Mangasarian class, and then prove that generated iteration points accumulate at bilevel KKT points under mild constraint qualifications. Our approach and analysis are applicable to a wide class of regularizers. Numerical comparisons demonstrate which smoothing functions work well for hyperparameter optimization via the bilevel optimization approach.

# Nonsmooth DC Optimization: Methods and Applications

Adil Bagirov
Institute of Innovation, Science and Sustainability
Federation University Australia
Ballarat, Victoria, Australia
E-mail: a.bagirov@federation.edu.au

**Abstract**. In this talk, we consider unconstrained and constrained optimization problems where the objective and/or constraint functions are represented as a difference of two convex (DC) functions. We discuss two different approaches to design methods of nonsmooth DC optimization: an approach based on the extension of bundle methods of nonsmooth optimization and an approach based on the DCA (difference of convex algorithm). We present numerical results and discuss various applications of DC optimization in machine learning.

# Optimization Models for Trend Filtering

Yoon Mo Jung
Department of Mathematics
Sungkyunkwan University, Korea
E-mail: yoonmojung@skku.edu

**Abstract**. Trend filtering is a regression problem to estimate underlying trends in time series data. It is necessary to investigate data in various disciplines. We propose a trend filtering method by adaptive piecewise polynomials. More specifically, we adjust the location and the number of breakpoints or knots to obtain a better fitting to given data. The numerical results on synthetic and real data sets show that it captures distinct features such as abrupt changes or kinks and provides a simplified form and brief summary of given data.

# Random Subspace Newton Method for Unconstrained Non-convex Optimization

Pierre-Louis Poirion
Continuous Optimization Team
Center for Advanced Intelligence Project
RIKEN, Japan
E-mail: pierre-louis.poirion@riken.jp

**Abstract**. In this talk, we present a randomized subspace regularized Newton method for a non-convex function. We show that our method has global convergence under appropriate assumptions, and its convergence rate is the same as that of the full regularized Newton method. Furthermore, we can obtain a local linear convergence rate, under some additional assumptions, and prove that this rate is the best we can hope when using random subspace.

# Active Neuron Least Squares: A Training Method for Multivariate Rectified Neural Networks

Yeonjong Shin
Department of Mathematical Sciences
KAIST, Korea
E-mail: yshin.math@gmail.com

**Abstract**. In this talk, we will present the Active Neuron Least Squares (ANLS), an efficient training algorithm for neural networks (NNs). ANLS is designed from the insight gained from the analysis of gradient descent training of NNs, particularly, the analysis of Plateau Phenomenon. The core mechanism is the option to perform the explicit adjustment of the activation pattern at each step, which is designed to enable a quick exit from a plateau. The performance of ANLS will be demonstrated and compared with existing popular methods in various learning tasks ranging from function approximation to solving PDEs.