Discrete Optimization and Machine Learning
RIKEN Center for Advanced Intelligence Project
Tokyo, Japan, 23-25 July, 2018

Organizers: Antoine Deza, Sebastian Pokutta, Takanori Maehara

ABSTRACT BOOKLET

Contents

Monday, July 23 **2**
- Session A1: 10h00-12h00 *Chair: Leo Liberti* . . . 2
- Session A2: 13h30-15h30 *Chair: Shinji Mizuno* . . . 3
- Session A3: 16h00-17h30 *Chair: Yao Xie* . . . 5

Tuesday, July 24 **6**
- Session B1: 10h00-12h00 *Chair: Yuri Faenza* . . . 6
- Session R1: 13h00-14h00 Robot Demo I . . . 7
- Session B2: 14h00-15h30 *Chair: Claudia D'Ambrosio* . . . 7
- Session B3: 16h00-17h30 *Chair: Sebastian Pokutta* . . . 8

Wednesday, July 25 **10**
- Session C1: 10h00-12h00 *Chair: Swati Gupta* . . . 10
- Session R2: 13h00-14h00 Robot Demo II . . . 11
- Session C2: 14h00-15h30 *Chair: Antoine Deza* . . . 12
- Session C3: 16h00-17h30 *Chair: Takanori Maehara* . . . 13

Monday, July 23

Session A1: 10h00-12h00 *Chair: Leo Liberti*

A11. **Blended conditional gradients**
Sebastian Pokutta (Georgia Tech)

We present a blended conditional gradient approach for minimizing a smooth convex function over a polytope P, that combines the Frank-Wolfe algorithm (also called conditional gradient) with gradient-based steps different from away steps and pairwise steps, however, still achieving linear convergence for strongly convex functions and good practical performance. Our approach retains all favorable properties of conditional gradient algorithms, most notably avoidance of projections onto P and maintenance of iterates as sparse convex combinations of a limited number of extreme points of P. The algorithm decreases measures of optimality (primal and dual gaps) rapidly, both in the number of iterations and in wall-clock time, outperforming even the efficient *lazified* conditional gradient algorithms of Braun et al. (see [arXiv:1410.8816]). Nota bene the algorithm is lazified itself. (joint work with Gábor Braun, Dan Tu, and Stephen Wright)

A12. **Projection-free simplex descent**
Dan Tu (Georgia Tech)

We present a new streamlined gradient descent algorithm for the probability simplex. This algorithm is at the core of the blended conditional gradient descent algorithm. In contrast to traditional projected gradient descent, we eschew projections completely but at the same time we overcome limitations of Frank-Wolfe algorithms. We achieve a linear per-iteration complexity and demonstrate in computational experiments that our method is superior to either projected gradient descent or Frank-Wolfe and we observe very significant computational speedups. We also achieve linear convergence for strongly convex functions. (joint work with Gábor Braun, Sebastian Pokutta, and Stephen Wright)

A13. **Finding new molecules using deep neural network, simulation, and search**
Kazuki Yoshizoe (Kazuki Yoshizoe)

Since chemical compounds could be described by atoms and bonds, novel molecule discovery is a combinatorial optimization problem. We tackle this problem by using AlphaGo-like approach by using RNN trained from chemical compound database, existing chemistry simulators, and Monte-Carlo Tree Search. Simplified Molecular-Input Line-Entry System (SMILES) is a popular way of describing molecules in ASCII string. By searching the space defined by SMILES using Monte-Carlo Tree Search (MCTS), we can find molecules which has a desired score. We use existing computational chemistry tools for calculating the physical properties of the molecules and use the physical values as the reward for MCTS. We obtained promising results as a Proof-of-Concept study by finding new molecules based on two different physical properties

A14. **Learning when to use a decomposition**
Axel Parmentier (Ecole Nationale des Ponts et Chaussśes)

Applying a Dantzig-Wolfe decomposition to a mixed-integer program (MIP) aims at exploiting an embedded model structure and can lead to significantly stronger reformulations of the MIP. Recently, automating the process and embedding it in standard MIP solvers have been proposed, with the detection of a decomposable model structure as key element. If the detected structure reflects the (usually unknown) actual structure of the MIP well, the solver may be much faster on the reformulated model than on the original. Otherwise, the solver may completely fail. We propose a supervised learning approach to decide whether or not a reformulation should be applied, and which decomposition to choose when several are possible. Preliminary experiments with a MIP solver equipped with this knowledge show a significant performance improvement on structured instances, with little deterioration on others. (joint work with Markus Kruber and Marco Lübbecke)

Session A2: 13h30-15h30 ***Chair: Shinji Mizuno***

A21. **Generative modeling via feasibility and iterative projection**
Naveen Kodali (Georgia Tech)

We consider the problem of modeling a probability distribution from a training set, and we focus on implicit generative models where one aims to directly learn a sampling oracle rather than a density function. We consider the training of such models through a hypothesis testing oracle: given any candidate distribution, we assume we can find potentially violated hypothesis tests, or report that no such hypotheses exist. We show that this framework, which has similarities to Generative Adversarial Networks, leads one to approach generative modeling as a feasibility problem rather than an optimization or min-max task, and that this perspective requires different analytical tools and also new estimation methods - in particular, we emphasize the use of iterative projection as the core algorithmic tool. We show that this alternative perspective exhibits both statistical benefits but also computational convergence guarantees.

A22. **Exact clustering via integer programming and maximum satisfiability**
Atsushi Miyauchi (RIKEN AIP)

We consider the following general graph clustering problem: given a complete undirected graph $G = (V, E, c)$ with an edge weight function $c : E \to \mathbb{Q}$, we are asked to find a partition $\mathcal{C}$ of V that maximizes the sum of edge weights within the clusters in $\mathcal{C}$. Owing to its high generality, this problem has a wide variety of real-world applications, including correlation clustering, group technology, and community detection. In this study, we investigate the design of mathematical programming formulations and constraint satisfaction formulations for the problem. First, we present a novel integer linear programming (ILP) formulation that has far fewer constraints than the standard ILP formulation by Grötschel and Wakabayashi (1989). Second, we propose an ILP-based exact algorithm that is usually faster than just solving our above ILP formulation. Third, we present maximum satisfiability (MaxSAT) counterparts of both our ILP formulation and ILP-based exact algorithm. Computational experiments demonstrate that our proposed approaches are highly effective in terms of both memory efficiency and computation time.

A23. **Approximation algorithms for covering problems**

Yotaro Takazawa (Tokyo Institute of Technology)

Covering problems are natural generalizations of the set covering problem. Since these problems are NP-hard, we study approximation algorithms as an approach to dealing with such problems. The goal of an approximation algorithm is to find an solution with a performance guarantee in polynomial-time. In this talk, we present approximation algorithms for two kinds of covering problems. First, we give primal-dual approximation algorithms for the covering $0-1$ integer program. Second, we propose LP-based approximation algorithms for partial covering problems, where some constraints may not be satisfied.

A24. **Submodular reassignment problem for reallocating agents to tasks with synergy effects**

Yoshio Okamoto (University of Electro Communications)

We propose a new combinatorial optimization problem that we call the submodular reassignment problem. We are given k submodular functions over the same ground set, and we want to find a set that minimizes the sum of the distances to the sets of minimizers of all functions. The problem is motivated by a two-stage stochastic optimization problem with recourse summarized as follows. We are given two tasks to be processed and want to assign a set of workers to maximize the sum of profits. However, we do not know the value functions exactly, but only know a finite number of possible scenarios. Our goal is to determine the first-stage allocation of workers to minimize the expected number of reallocated workers after a scenario is realized at the second stage. This problem can be modeled by the submodular reassignment problem. We prove that the submodular reassignment problem can be solved in strongly polynomial time via submodular function minimization. We further provide a maximum-flow formulation of the problem that enables us to solve the problem without using a general submodular function minimization algorithm, and more efficiently both in theory and in practice. In our algorithm, we make use of Birkhoff's representation theorem for distributive lattices. (joint work with Naoyuki Kamiyama, Naonori Kakimura, and Yusuke Kobayashi)

Session A3: 16h00-17h30 ***Chair: Yao Xie***

A31. **Deep reinforcement learning in inventory management**
Jana Boerger (Georgia Tech)

Inventory Control problems require an agent to make decisions on the order quantity from a supplier to meet future customer demand. Traditionally, these problems have been solved optimally by solving the Bellman Equations for the related Markov Decision Process (MDP) to derive an (s, S)-policy. Reinforcement Learning (RL) approaches have been applied in the past but the state space has been too high to prove to be viable. We show that using a Deep Q-Network (DQN) the optimal (s, S)-policy can be learned and furthermore DQN can outperform the traditional approach in more complicated settings.

A32. **A framework of efficient online learning algorithms and applications**
Nguyen Kim Thang (Université d'Evry)

We consider the design of oracle-efficient online learning algorithms. In the setting, we have access to an offline oracle and the goal is to design online learning algorithms that at any step return a decision by making a polynomial number of oracle calls and achieve the total gain approximately close to that obtained by the best fixed decision in hindsight. It has been shown that such online algorithms do not exist in general adversarial environments and a major challenge is to characterize the structural regularities of specific settings under which oracle-efficient online learning algorithms exist. In the paper, we present a framework and sufficient conditions to design oracle-efficient online learning algorithms. We show the applicability of our framework to mechanism design and bandit problems.

A33. **Machine learning from weak supervision**
Masashi Sugiyama (University of Tokyo / RIKEN AIP)

Machine learning from big labeled data is highly successful in speech recognition, image understanding and natural language translation. On the other hand, there are various application domains where human labor is involved in the data acquisition process and thus the use of massive labeled data is prohibited. In this talk, I will introduce our recent advances in machine learning techniques from limited supervision.

Tuesday, July 24

Session B1: 10h00-12h00 ***Chair: Yuri Faenza***

B11. **Discrete *DC* programming for integrally convex functions**
Kazuo Murota (Tokyo Metropolitan University)

A theoretical framework of discrete *DC* programming was proposed by Maehara and Murota (Mathematical Programming, 2015), with particular reference to M-convex and L-convex functions in discrete convex analysis. In this talk we extend this framework to integrally convex functions, by establishing integral biconjugacy for integer-valued integrally convex functions. In particular, the discrete Toland-Singer duality is extended to integrally convex functions.

B12. **On limited memory Kelley's method**
Swati Gupta (Georgia Tech)

The original simplicial method (OSM), a variant of the classic Kelley's cutting plane method, has been shown to converge to the minimizer of composite convex and submodular objectives, though no rate of convergence for this method was known. Moreover, OSM is required to solve subproblems in each iteration whose size grows linearly in the number of iterations. We propose a limited memory version of Kelley's method (L-KM) that is a novel adaptation of OSM and requires limited memory independent of the iteration (at most $n+1$ constraints for an n-dimensional problem), while maintaining convergence to the optimal solution. We further show that the dual method of L-KM is a special case of the Fully-Corrective Frank-Wolfe (FCFW) method with approximate correction, thereby deriving a limited memory version of FCFW method and proving a rate of convergence for L-KM. Though we propose L-KM for minimizing composite convex and submodular objectives, our results on limited memory version of FCFW hold for general polytopes, which is of independent interest.

B13. **Submodular maximization over logics**
Takanori Maehara (RIKEN AIP)

Maximizing a monotone submodular function over some constraint is a fundamental problem in discrete optimization, and has many applications in machine learning. The problem cannot be solved exactly even for the cardinality constraint; however, it admits efficient provable approximation algorithms for several constraints such as cardinality constraint, knapsack constraint, and matroid constraint. In this study, we consider the case that the constraint represents subgraphs of a graph specified by some logic formula. We show that there is $n^{O(1)}$ time $\log n$ approximation algorithm if the graph has bounded treewidth and the logic is in MSO, and $O(n^{log\, k})$ time $\log k$ approximation algorithm if the graph has bounded expansion and the logic is in FO. (joint work with Thomas Rigoux and Masakazu Ishihata)

B14. **Finding Submodularity hidden in symmetric difference**

Shuji Kijima (Kyushu University)

Submodular functions are often compared with convex functions, as a discrete analogy. In this talk, we are concerned with symmetric difference transformations (SD-transformations) of submodular functions, as the counter part of affine transformations of convex functions. It is a well known fact that convexity is invariant under affine transformations. In contrast, submodularity is not preserved under SD-transformations. This talk presents a characterization of SD-transformations preserving submodularity. Then, we discuss the problem of discovering a *canonical set* S, given the SD-transformation g of a submodular function f by S, provided that $g(X)$ is given by a function value oracle. (joint work with Junpei Nakashima, Yukiko Yamauchi and Masafumi Yamashita)

Session R1: 13h00-14h00 **Robot Demo I**

Cognitive Assistive Technology team
Mihoko Otake (University of Tokyo / RIKEN AIP)

Session B2: 14h00-15h30 ***Chair: Claudia D'Ambrosio***

B21. **Random projections in mathematical programming**
Leo Liberti (CNRS and Ecole Polytechnique)

In the algorithmic trade-off between generality and efficiency, sometimes the only way out is to accept approximate methods. If all else fails, we can always fall back on heuristic methods, but some form of approximation guarantee is usually preferable. In this talk we shall discuss a set of approximating reformulations to various classes of mathematical programming problems involving matrices. Random projections are a dimensionality reduction methodology which projects a set of vectors into a much lower dimensional space, so that the projected vector set is approximately congruent to the original one with high probability. The probability of failure falls exponentially fast as the dimension increases, making this a truly *big data* methodology. We shall show how to apply this methodology to linear, conic and (bounded) quadratic programming.

B22. **An iterative algorithm using random projection for LP**
Pierre-Louis Poirion (CNRS and Ecole Polytechnique)

Random projections allow to approximate a Linear Program, with many constraints, by forming a new LP, with much fewer constraints, in randomly aggregating them. We will present a new algorithm that iterates this process and converge to an optimal solution,

B23. **Randomized Sketches for feasibility of large linear systems subject to arbitrary convex constraints**
Khac Ky Vu (Chinese University of Hong Kong)

Randomized sketches provide a simple but very powerful technique that used for reducing dimensions. In this talk, I will apply randomized sketches to the problem of establishing the feasibility of a linear system subject to arbitrary convex constraints. In particular, we test the feasibility by sketching the linear system to obtain a simpler system while keeping the convex constraints unchanged. In the case the feasibility is validated, we propose a sketch-and-project algorithm to obtain a feasible solution. Our results can be applied in contexts when the access to the original data is limited or unavailable (e.g. preservation of privacy).

Session B3: 16h00-17h30 ***Chair: Sebastian Pokutta***

B31. **Fenchel-Moreau conjugation inequalities in generalized convexity**
Michel De Lara (Ecole Nationale des Ponts et Chaussées)

There are many optimization problems where one minimizes the sum of two functions, with one of them depending on parameters: for instance in machine learning (Lasso) and, more generally in optimization (perturbation of constraints). We provide Fenchel-Moreau conjugation inequalities for a class of such minimization problems, but with general couplings. We display several applications. We provide a new formula for the Fenchel-Moreau conjugate of a generalized inf-convolution. We obtain formulas with partial Fenchel-Moreau conjugates. Finally, we consider the Bellman equation in stochastic dynamic programming and we provide a *Bellman-like* equation for the Fenchel conjugates of the value functions.

B32. **Heterogeneous medical data analysis by selective inference**
Ichiro Takeuchi (Nagoya Institute of Technology / RIKEN AIP)

In this talk, we introduce Selective Inference, which has been recently demonstrated to be an effective statistical analysis technique for data-driven scientific discovery, and mainly talk about its application to heterogeneous medical data analysis. When analyzing a medical data set containing heterogeneous samples, one commonly taken approach is the following two-stage method: (1) identify homogeneous groups of samples by clustering, and (2) find genetic factors which characterize each of the identified groups. Unfortunately, however, such two-stage analysis results are often biased since the clustering in the 1st stage and the inference in the 2nd stage are conducted by using same data. In this talk, we show that, by using selective inference framework, we can effectively correct the clustering bias, and demonstrate its effectiveness by applying the method to single cell data analysis.

B33. **Fast yet simple natural-gradient variational inference in complex models**
Mohammad Emtiyaz Khan (RIKEN AIP)

Approximate Bayesian inference is promising in improving generalization and reliability of deep learning, but is computationally challenging. Modern variational-inference (VI) methods circumvent the challenge by formulating Bayesian inference as an optimization problem and then solving it using gradient-based methods. In this talk, I will argue in favor of natural-gradient approaches which can improve convergence of VI by exploiting the information geometry of the solutions. I will discuss a fast yet simple natural-gradient method obtained by using a duality associated with exponential-family distributions. I will summarize some of our recent results on Bayesian deep learning, where natural-gradient methods lead to an approach which gives simpler updates than existing VI methods while performing comparably to them.

Wednesday, July 25

Session C1: 10h00-12h00 ***Chair: Swati Gupta***

C11. **Robust hypothesis testing using Wasserstein uncertainty**
Yao Xie (Georgia Tech)

We develop a novel computationally efficient and general framework for robust hypothesis testing. The new framework features a new way to construct uncertainty sets under the null and the alternative distributions, which are sets centered around the empirical distribution defined via Wasserstein metric, thus our approach is data-driven and free of distributional assumptions. We develop a convex safe approximation of the minimax formulation and show that such approximation renders a nearly-optimal detector among the family of all possible tests. By exploiting the structure of the least favorable distribution, we also develop a tractable reformulation of such approximation, with complexity independent of the dimension of observation space and can be nearly sample-size-independent in general. Real-data example using human activity data demonstrated the excellent performance of the new robust detector. (joint work with Rui Gao, Liyan Xie, and Huan Xu)

C12. **Statistical estimation for non-smooth functions with the regularity lemma**
Masaaki Imaizumi (Institute of Statistical Mathematics / RIKEN AIP)

We consider estimation methods for non-smooth functions using the regularity lemma. Estimation for functions appears in various situations, and it is intensively used in statistics and machine learning. The statistical efficiencies of estimators, i.e., their convergence rates, play a central role in advanced statistical analysis. Although estimators and their convergence rates for smooth functions are well investigated in the literature, those for non-smooth density functions remain elusive despite their importance in application fields. In this paper, we propose new techniques for non-smooth functions by employing the notion of Szemerédi partitions from graph theory. We investigate their properties including their convergence rates.

C13. **Semiparametric choice models**
Selin Ahipasaoglu (Singapore University of Technology and Design)

Discrete choice models are widely used in the analysis of individual choice behaviour in many fields, including economics, transportation systems, manufacturing, and marketing. They can also be used for clustering and categorization in machine learning applications. Most existing discrete choice models use the Random Utility Model (RUM), which is characterized by an explicit assumption on the distribution of the random utilities of the discrete alternatives. Examples include the multinomial logit (MNL), nested logit (NL), and multinomial probit (MNP) models. However, except in a few cases, the MNL variants suffer from undesirable properties, such as the Independence of Irrelevant Alternatives (IIA) property, since they lack the ability to fully incorporate the correlation information between alternatives.

The MNP model on the other hand is flexible and can make good use of correlations. Nevertheless, calculating the choice probabilities requires computationally intensive simulation methods, making the MNP choice model intractable for any realistic application. We present alternative discrete choice models that relax the Random Utility Model (RUM) assumption, but use a distributionally robust framework to model utilities. In particular, they assume that a system planner, whose goal is to study, estimate, and predict the choice behaviour, is not willing to estimate the full joint distribution of the random utility terms. Instead, he estimates partial information, such as moments or marginals, and constructs a model which searches over all possible distributions that fit to the assumed partial distribution. The choice probabilities are evaluated under an extremal probability distribution which maximizes the utility of a representative decision maker given partial information. These models are referred to as *semiparametric choice models* and have important advantages in terms of estimation, modelling flexibility, and computational tractability, which are useful for high dimensional applications.

C14. **Mathematical programming for limited memory influence diagram**
Victor Cohen (Ecole Nationale des Ponts et Chausssées)

Probabilistic graphical models are a tool of the Machine Learning community to deal with high dimensional probability distributions. Limited Memory Influence Diagrams (LIMID) leverage probabilistic graphical models to represent discrete stochastic optimization problems, including Markov Decision Process (MDP) and Partially Observable MDP as standard examples. More precisely, given random variables considered as vertices of an acyclic digraph, a probabilistic graphical model defines a joint distribution via the conditional distributions of vertices given their parents. In LIMIDs, the random variables are represented by a probabilistic graphical model whose vertices are partitioned into three types : chance, decision and utility vertices. The user chooses the distribution of the decision vertices conditionally to their parents in order to maximize the expected utility. Leveraging the notion of strong junction tree, we present a mixed integer linear formulation for solving a LIMID, as well as valid inequalities, which leads to a computationally efficient algorithm. We also show that the linear relaxation yields an optimal integer solution for instances that can be solved by the *single policy update*, the default algorithm for addressing LIMIDs.

Session R2: 13h00-14h00 **Robot Demo II**

Hakuto, the Japanese Google-X team
Takeshi Hakamada (Ispace)

Session C2: 14h00-15h30 *Chair: Antoine Deza*

C21. **Legal assignments and fast EADAM with consent via classical theory of stable matchings**
Yuri Faenza (Columbia University)

Gale and Shapley's college admission problem has been extensively studied, applied, and extended in the last decades. The input of the classical model is a set of students and schools, with students having a strictly ordered list of schools they would like to attend, and similarly schools having a strictly ordered list of students they are willing to admit, as well as the maximum number of seats available at each school (quota). The output is an assignment of students to schools that respects quotas and a concept of fairness known as stability. The goal of many extensions of this model is to obtain an assignment more favorable to students, often shifting the focus from stability to (constrained) Pareto efficiency (for students). In this talk, we show how an improved structural and algorithmic understanding of some of those extensions can be obtained reducing those problems back to the original stable matching problem, and exploiting / extending classical theory developed therein. (joint work with Xuan Zhang)

C22. **Tensor factorization for spatio-temporal predictions**
Koh Takeuchi (NTT CS Labs)

Analysis of spatio-temporal data is a common research topic that requires the interpolations of unknown locations and the predictions of feature observations by utilizing information about where and when the data were observed. One of the most difficult problems is to make predictions of unknown locations. Tensor factorization methods are popular in this field because of their capability of handling multiple types of spatio-temporal data, dealing with missing values, and providing computationally efficient parameter estimation procedures. However, unlike traditional approaches such as spatial autoregressive models, the existing tensor factorization methods have not tried to learn spatial autocorrelations. These methods employ previously inferred spatial dependencies, often resulting in poor performances on the problem of making interpolations and predictions of unknown locations. In this paper, we propose a new tensor factorization method that estimates low-rank latent factors by simultaneously learning the spatial and temporal autocorrelations. We introduce new spatial autoregressive regularizers based on existing spatial autoregressive models and provide an efficient estimation procedure. With experiments on publicly available traffic transporting data, we demonstrate that our proposed method significantly improves the predictive performances in our problems in comparison to the existing state-of-the-art spatio-temporal analysis methods.

C23. **Using linear programming to train neural networks**
Gonzalo Muñoz (Polytechnique Montréal)

Deep Learning received significant attention due to its impressive performance in many state-of-the-art learning tasks. Unfortunately, while very powerful, Deep Learning is not well understood and in particular, only recently results for the complexity of training deep neural networks have been obtained. In this work, we show that large classes of deep neural networks with various architectures, activation functions, and loss functions can be trained to near optimality with a desired target accuracy using linear programming. The size of this LP is exponential in the size of the architecture and linear in the size of the data, as opposed to best known exact results which have an exponential dependency in the data. Our results shed new light in the overall hardness of generic empirical risk minimization problems which can be used to address previously unstudied architectures in learning problems.

Session C3: 16h00-17h30 ***Chair: Takanori Maehara***

C31. **Spectral sparsification of hypergraph**

Yuichi Yoshida (National Institute of Informatics)

For an undirected/directed hypergraph $G = (V, E)$, its Laplacian $L_G\colon \mathbb{R}^V \to \mathbb{R}^V$ is defined such that its *quadratic form* $x^\top L_G(x)$ captures the cut information of G. In particular, $\mathbf{1}_S^\top L_G(\mathbf{1}_S)$ coincides with the cut size of $S \subseteq V$, where $\mathbf{1}_S \in \mathbb{R}^V$ is the characteristic vector of S. A weighted subgraph H of a hypergraph G on a vertex set V is said to be an ϵ-spectral sparsifier of G if $(1-\epsilon)x^\top L_H(x) \leq x^\top L_G(x) \leq (1+\epsilon)x^\top L_H(x)$ holds for every $x \in \mathbb{R}^V$. In this paper, we present a polynomial-time algorithm that, given an undirected/directed hypergraph G on n vertices, constructs an ϵ-spectral sparsifier of G with $O(n^3 \log n/\epsilon^2)$ hyperedges/hyperarcs. The proposed spectral sparsification can be used to improve the time and space complexities of algorithms for solving problems that involve the quadratic form, such as computing the eigenvalues of L_G, computing the effective resistance between a pair of vertices in G, semi-supervised learning based on L_G, and cut problems on G. In addition, our sparsification result implies that any submodular function $f\colon 2^V \to \mathbb{R}_+$ with $f(\emptyset) = f(V) = 0$ can be concisely represented by a directed hypergraph. Accordingly, we show that, for any distribution, we can properly and agnostically learn submodular functions $f\colon 2^V \to [0,1]$ with $f(\emptyset) = f(V) = 0$, with $O(n^4 \log(n/\epsilon)/\epsilon^4)$ samples.

C32. **Accelerated stochastic optimization for finite sum regularized empirical risk minimization**

Taiji Suzuki (University of Tokyo / RIKEN AIP)

Covering problems are natural generalizations of the set covering problem. Since these problems are NP-hard, we study approximation algorithms as an approach to dealing with such problems. The goal of an approximation algorithm is to find an solution with a performance guarantee in polynomial-time. In this talk, we present approximation algorithms for two kinds of covering problems. First, we give primal-dual approximation algorithms for the

covering $0-1$ integer program. Second, we propose LP-based approximation algorithms for partial covering problems, where some constraints may not be satisfied.

C33. **Convex matroid optimization, lattice polytopes, and degree sequences of hypergraphs**

Antoine Deza (McMaster University)

We introduce a family of polytopes, called primitive zonotopes, which can be seen as a generalization of the permutahedron of type B_d. We discuss connections to the largest diameter of lattice polytopes and to the computational complexity of multicriteria matroid optimization. Complexity results and open questions are also presented. In particular, we answer a question raised in 1986 by Colbourn, Kocay, and Stinson by showing that deciding whether a given sequence is the degree sequence of a 3-hypergraph is computationally prohibitive. (joint work with Asaf Levin, George Manoussakis, Syed Meesum, Shmuel Onn and Lionel Pounin)

Speaker index

Ahipasaoglu, Selin, 10

Boerger, Jana, 5

Cohen, Victor, 11

De Lara, Michel, 8
Deza, Antoine, 14

Faenza, Yuri, 12

Gupta, Swati, 6

Hakamada, Takeshi, 11

Imaizumi, Masaaki, 10

Khan, Mohammad Emtiyaz , 9
Kijima, Shuji, 7
Kodali, Naveen, 3

Liberti, Leo, 7

Maehara, Takanori, 6
Miyauchi, Atsushi, 3
Muñoz, Gonzalo, 12
Murota, Kazuo, 6

Okamoto, Yoshio, 4
Otake, Mihoko, 7

Parmentier, Axel, 3
Poirion, Pierre-Louis , 7
Pokutta, Sebastian, 2

Sugiyama, Masashi, 5
Suzuki, Taiji, 13

Takazawa, Yotaro, 4
Takeuchi, Ichiro, 8
Takeuchi, Koh, 12
Thang, Nguyen Kim, 5
Tu, Dan, 2

Vu, Khac Ky, 8

Xie, Yao, 10

Yoshida, Yuichi, 13
Yoshizoe, Kazuki, 2