The First International Conference on Mathematics of Data Science (MathDS)

20-24 March 2017

Organizers:

Department of Mathematics of Hong Kong Baptist University
The Liu Bie Ju Centre for Mathematical Sciences of City University of Hong Kong

Organizing Committee:

Charles K. Chui (HKBU)
Michael Ng (HKBU)
Ding-Xuan Zhou (CityU)

Program

20 Mar 2017 (Mon) SWT501, Shaw Tower, Shaw Campus

Morning Session Chair: Michael Ng

09:20-09:30	Opening
09:30-10:10	Charles Chui *Means for knowledge dissemination with samples and case studies*
10:10-10:50	Gerlind Plonka *Sparse phase retrieval of one-dimensional signals by Prony's method*
10:50-11:10	Break
11:10-11:50	Sergei Pereverzyev *Balancing principle in supervised learning for a general regularization scheme*
11:50-12:30	Dao-Qing Dai *Sparse dictionary regression with applications*
12:30-14:00	Lunch (by invitation)

Afternoon Session Chair: Charles Chui

14:00-14:40	Matthew Hirn *Learning many body physics via multiscale, multilayer machine learning architectures*
14:40-15:20	Jianzhong Wang *Ensemble boosting algorithms based on data sorting for semi-supervised classification*
15:20-16:00	Break
16:00-16:40	Tao Qian *Adaptive orthonormal systems for matrix-valued functions*
16:40-17:20	M.D. van der Walt *Atomic signal decomposition via SuperEMD*
17:20-18:00	Haizhao Yang *Recursive schemes for shape functions in the mode decomposition problem*

21 Mar 2017 (Tue) SWT501, Shaw Tower, Shaw Campus

Morning Session Chair: Ding-Xuan Zhou

09:30-10:10	Frèdèric Chazal *Data driven estimation of the Laplace-Beltrami operator*
10:10-10:50	Raymond H. Chan *Geometric tight frame based stylometry for an Authentication of van Gogh paintings*
10:50-11:10	Break
11:10-11:50	Zhiqiang Xu *The minimal measurement number for low-rank matrix recovery*
11:50-12:30	Jianfeng Cai *Non-convex methods for low-rank matrix reconstruction*
12:30-14:00	Lunch (by invitation)

Afternoon Session Chair: Henry Ngan

14:00-14:40	Wenchang Sun *Frames of uniform subframe bounds with applications to erasures*
14:40-15:20	Shaobo Lin *Distributed semi-supervised learning*
15:20-16:00	Break
16:00-16:40	Ting Hu *Convergence of gradient descent for kernel minimum error entropy principle*
16:40-17:20	Junhui Wang *Gradient-induced model-free variable selection*
17:20-18:00	Can Yang *Adaptive False Discovery Rate regression with application in integrative analysis of large-scale genomic data*

22 Mar 2017 (Wed) SWT501, Shaw Tower, Shaw Campus

Morning Session Chair: Charles Chui

09:30-10:10	Yuesheng Xu *Mathematics in data science*
10:10-10:50	Sergiy Pereverzyev Jr. *Adaptive Nyström subsampling for dealing with Big Data*
10:50-11:10	Break
11:10-11:50	Lihua Yang *Questions on mono-components*
11:50-12:30	Xin Guo *Convergence of the Randomized Kaczmarz Algorithm in Hilbert Space*
12:30-14:00	Lunch (by invitation)

Afternoon Session Chair: Michael Ng

14:00-14:40	Xiaoming Huo *Challenges in data science and a possible roadmap of future work*
14:40-15:20	Qiang Wu *Bias correction for regularized kernel regression with applications in distributed learning*
15:20-16:00	Break
16:00-16:40	Weiguo Gao *Algorithms for group sparsity with overlap and beyond*
16:40-17:20	Wai-Ki Ching *Construction of probabilistic boolean networks with applications*
17:20-18:00	Xiaosheng Zhuang *Multiscale data analysis: Framelets, manifolds and graphs*
18:30	Banquet (by invitation)

23 Mar 2017 (Thu) FSC501, Fong Shu Chuen Library, Ho Sin Hang Campus

Morning Session Chair: Henry Ngan

09:30-10:10	Simon Foucart *Computing a quantity of interest from observational data*
10:10-10:50	Lek-Heng Lim *Tensor network ranks*
10:50-11:10	Break
11:10-11:50	Jian Huang *A constructive approach to L0-penalized linear regression*
11:50-12:30	Hongzhi Tong *Calibration of ε-insensitive loss in support vector machines regression*
12:30-14:00	Lunch (by invitation)

24 Mar 2017 (Fri) SWT501, Shaw Tower, Shaw Campus

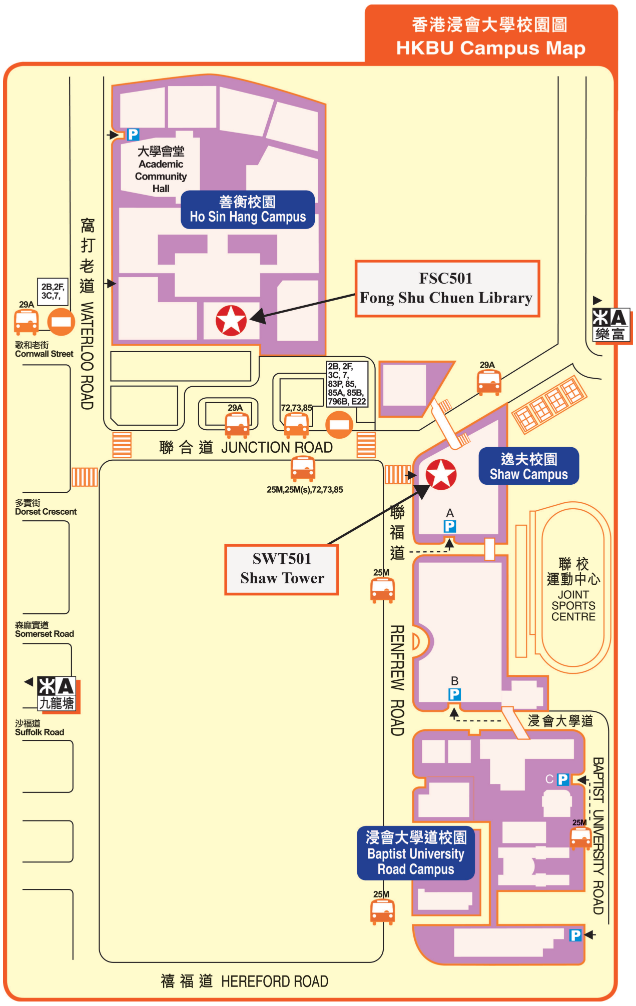
Morning Session Chair: Henry Ngan

09:30-10:10	Holger Wendland *Kernel-based reconstructions for parametric PDEs*
10:10-10:50	Ronald Lok Ming Lui *Restoration of atmospheric turbulence-distorted images via RPCA and Quasiconformal maps*
10:50-11:10	Break
11:10-11:50	Sou-Cheng Terrya Choi *Probabilistic record linkage and address standardization*
11:50-12:30	Xiaojun Chen *Nonsmooth, nonconvex regularization for sparse optimization*
12:30-14:00	Lunch (by invitation)

Afternoon Session Chair: Ding-Xuan Zhou

14:00-14:40	Xiaoming Yuan *How to implement ADMM to large-scale datasets?*
14:40-15:20	Yingchun Jiang *Spatially distributed sampling and reconstruction of signals on a graph*
15:20-16:00	Yuan Yao *Boosting with structural sparsity – A differential inclusion approach*
	Closing

Campus Map



Transportation

(i) Shuttle bus from Harbour Plaza Metropolis to HKBU
pick-up time
20 March 8:40 a.m.
21-24 March 8:50 a.m.
meeting point: hotel entrance

(ii) Take MTR from Hung Hum Station to Kowloon Tong Station and Walk to HKBU Campus

Abstracts

Non-convex methods for low-rank matrix reconstruction

Jianfeng Cai
Hong Kong University of Science and Technology, Hong Kong
jfcai@ust.hk

We present a framework of non-convex methods for reconstructing a low rank matrix from its limited information, which arises from numerous practical applications in machine learning, imaging, signal processing, computer vision, etc. Our methods will be applied to several concrete example problems such as matrix completion, phase retrieval, and spectral compressed sensing with super resolution. We will also provide theoretical guarantee of our methods for the convergence to the correct low-rank matrix.

Geometric tight frame based stylometry for an Authentication of van Gogh paintings

Haixia Liu, Raymond H. Chan*, and Yuan Yao
Chinese University of Hong Kong, Hong Kong
rchan@math.cuhk.edu.hk

In this talk, we propose a new stylometric analysis based on a geometric tight frame and some low order moment statistics which are used to extract features from the paintings. A small set of these features are selected by a forward stage-wise rank boosting so that van Gogh paintings are highly concentrated towards some center point and forgeries are spread out as outliers. In numerical results, we can achieve 86.1% classification accuracy under leave-one-out cross-validation. With only 5 selected features, our method can give 88.6% classification accuracy. Our work shows that a small set of statistics of the tight-frame coefficients along certain orientations can serve as discriminative features for van Gogh paintings. It reflects a highly consistent style in van Goghs brushstroke movements, where many forgeries demonstrate a more diverse spread in these features.

Research is supported by HKRGC.

Data driven estimation of the Laplace-Beltrami operator

Frédéric Chazal
Inria Saclay, France
frederic.chazal@inria.fr

Approximations of Laplace-Beltrami operators on manifolds through graph Laplacians have become popular tools in data analysis and machine learning. These discretized operators usually depend on bandwidth parameters whose tuning remains a theoretical and

practical problem. In this talk, we address this problem for the unnormalized graph Laplacian by establishing an oracle inequality that opens the door to a well-founded data-driven procedure for the bandwidth selection. Our approach relies on recent results by Lacour and Massart on the so-called Lepskis method.

This is a joint work with Ilaria Giulini (Inria Saclay & Universit Paris Diderot) and Bertrand Michel (Ecole Centrale de Nantes)

Nonsmooth, nonconvex regularization for sparse optimization

Xiaojun Chen
The Hong Kong Polytechnic University, Hong Kong
xiaojun.chen@polyu.edu.hk

We consider a class of constrained minimization problems where the objective function is a sum of a smooth function and a nonsmooth, nonconvex, perhaps even non-Lipschitz regularization. On concave regularization including SCAD, MCP and L-p norm ($0 < p < 1$), we show that finding a global optimal solution is strongly NP-hard. On the other hand, we present lower bounds of nonzero entries in every local optimal solution. Such lower bounds can be used to classify zero and nonzero entries in local optimal solutions and select regularization parameters for desirable sparsity of local optimal solutions. Moreover, we introduce several efficient algorithms including smoothing quadratic regularization algorithms, smoothing trust region Newton methods, interior point algorithms and augmented Lagrangian methods. Examples of sparse portfolio selection are presented to illustrate the theory and algorithms.

Construction of probabilistic boolean networks with applications

Wai-Ki Ching
The University of Hong Kong, Hong Kong
wching@hku.hk

Boolean Networks (BNs) and their extensions Probabilistic Boolean Networks (PBNs) are useful models for studying genetic regulatory interactions and many other real-world problems. For the purpose of network inference and system synthesis, one has to construct such a network. It is a challenging problem, because there may be many networks or no network having the required properties. The construction of PBNs from observed data sets (e.g. a given transition-probability matrix) is an interesting problem of huge size. In this talk, we shall propose a construction method with some applications. Numerical examples will be given to demonstrate the effectiveness of the proposed method.

Probabilistic record linkage and address standardization

Sou-Cheng Terrya Choi
Department of Applied Mathematics, Illinois Institute of Technology, USA
schoi32@iit.edu

Probabilistic record linkage (PRL) refers to the process of matching records from various data sources such as database tables with some missing or corrupted index values. Human is often involved in a loop to review cases that the algorithm cannot match. PRL can be applied to join or de-duplicate records, or to impute missing data, resulting in better overall data quality. An important subproblem in PRL is to parse a field such as address into its components such as street number, street name, city, state, and zip code. Various data analysis techniques such as natural language processing and machine learning methods are often gainfully employed in both PRL and address standardization to achieve higher accuracies of linking or prediction. This work compares the performance of four reputable PRL packages freely available in the public domain, namely FRIL, Link Plus, R RecordLinkage, and SERF. In addition, we evaluate the baseline performance and sensitivity of four address-parsing web services including the Data Science Toolkit, Geocoder.us, Google Maps APIs, and the U.S. address parser. Finally, we present some of the strengths and limitations of the software and services we have evaluated. This is joint work with Yongheng Lin and Edward Mulrow.

Means for knowledge dissemination with samples and case studies

Charles Chui
Hong Kong Baptist University, Hong Kong

"Data Science", an interdisciplinary research area, originated within Computer Science and a major subject of studies in Statistics, has recently made a major media buzz, not only in many industrial sectors, but also among academic institutions, world wide. In the high-tech and financial industries, demand has outpace the supply of data scientists, with bigger paychecks over analytics professionals and software developers. Many top-tier universities have been hiring new Ph.D.'s and young post-docs with proven adequate training and research experience in massive/complex data analysis and high potential to advancing this high-demand interdisciplinary field. On the other hand, as we all know, mathematicians have been developing rigorous theory, as well as powerful computational methods and effective algorithms, throughout the past two thousand years, to solving many Big Data problems; such as the elegant system identification algorithm, introduced in 1795 by Gaspard de Prony (1755 -1839), and the ingenious least-square method based on the conic section model for computing planet orbits, by Carl Friedrich Gauss (1777-1855), for the successful rediscovery of the dwarf planet Ceres in October, 1801. Current successful mathematical development relevant to the advancement of Data Science, such as the research findings to be presented in this conference, includes but is not limited to: deep learning and deep net (neural network) realization, sparse data representation, compressed sensing, compressive sampling, harmonic analysis on graphs, machine and manifold learning, and blindsource data separation. However, it is unfortunate for the current scientific and technological advancement that the current means for knowledge dissemination of mathematical discovery are not easily accessible to the data science users. In fact, data science research journals have been restricted to applications to Business, e-Commerce, Economics, Computer Science, Geoscience, Life sciences, Social Science, Health Care, Environment and Climate, etc. In this presentation, we will discuss more effective means for dissemination of mathematical advancement relevant to data science, by announcing a newly founded open-access journal, with samples of published papers and case studies of somewhat non-traditional means for knowledge dissemination of Mathematics of Data Science.

Sparse dictionary regression with applications

Dao-Qing Dai
Sun Yat-sen University, China
stsddq@mail.sysu.edu.cn

Regression with dictionary in high dimensional setting is attracting attention recently. In this talk I shall report constructions of discriminant dictionary and their applications to face recognition.

Computing a quantity of interest from observational data

Simon Foucart
Texas A&M University, USA
foucart@tamu.edu

Scientific problems often feature observational data received in the form $w_1 = l_1(f), \ldots$, $w_m = l_m(f)$ of known linear functionals applied to an unknown function f from some Banach space $\mathcal{X}$, and it is required to either approximate f (the full approximation problem) or to estimate a quantity of interest $Q(f)$. In typical examples, the quantities of interest can be the maximum/minimum of f or some averaged quantity such as the integral of f, while the observational data consists of point evaluations. To obtain meaningful results about such problems, it is necessary to possess additional information about f, usually as an assumption that f belongs to a certain model class $\mathcal{K}$ contained in $\mathcal{X}$. This is precisely the framework of optimal recovery, which produced substantial investigations when the model class is a ball of a smoothness space, e.g. when it is a Lipschitz, Sobolev, or Besov class. This presentation is concerned with other model classes described by approximation processes. Its main contributions are: (i) for the estimation of quantities of interest, the production of numerically implementable algorithms which are optimal over these model classes, (ii) for the full approximation problem, the construction of linear algorithms which are optimal or near optimal over these model classes in case of data consisting of point evaluations. Regarding (i), when Q is a linear functional, the existence of linear optimal algorithms was established by Smolyak, but the proof was not numerically constructive. In classical recovery settings, it is show here that such linear optimal algorithms can be produced by constrained minimization methods, and examples involving the computations of integrals from the given data are examined in greater details. Regarding (ii), it is shown that linearization of optimal algorithms can be achieved for the full approximation problem, too, in the important situation where the l_j are point evaluations and $\mathcal{X}$ is a space of continuous functions equipped with the uniform norm. It is also revealed how the quasi-interpolation theory allows for the construction of linear algorithms which are near optimal.

Algorithms for group sparsity with overlap and beyond

Weiguo Gao
Fudan University, China
wggao@fudan.edu.cn

We discuss the algorithms for group sparsity with overlap in which matrix-vector multiplications can be carried out efficiently. An accelerating technique to improve the performance of existing algorithms is proposed by solving a similar sub-problem. We further show that the new scheme can be formulated in the inner-outer regime by exploring the sparse structure. This can also be viewed from the dimensional reduction viewpoint. Convergence for ADMM (as an example) is guaranteed by a rigorous proof. Numerical experiments demonstrate the efficiency of the technique.

Convergence of the Randomized Kaczmarz Algorithm in Hilbert Space

Xin Guo
The Hong Kong Polytechnic University, Hong Kong
x.guo@polyu.edu.hk

The randomized Kaczmarz algorithm recently draws much attention. Existing results on analysis suffer from condition numbers of the linear equation systems. Although the randomized Kaczmarz algorithm has a natural generalization to Hilbert spaces (which covers online learning algorithms for a particular instance), the existing analysis does not. Although the large-scale linear equation system is an ideal scenario for the randomized Kaczmarz algorithm to outperform direct solvers, it is also a scenario the existing analysis is not satisfactory. In this research, we introduce the regularity assumption widely adopted in learning theory and obtain the polynomial convergence rate of the randomized Kaczmarz algorithm in Hilbert space under noise-free setting. We find that by nature, the randomized Kaczmarz algorithm converges weakly. Meanwhile, with noisy data, we study the relaxation method and obtain a strong convergence arbitrarily close to the minimax optimal rate. The result applies to online gradient descent learning algorithms and significantly improves the existing learning rate in literature.

Learning many body physics via multiscale, multilayer machine learning architectures

Matthew Hirn
Michigan State University, USA
mhirn@msu.edu

Deep learning algorithms are making their mark in machine learning, obtaining state of the art results across computer vision, natural language processing, auditory signal processing and more. A wavelet scattering transform has the general architecture of a convolutional neural network, but leverages structure within data by encoding multiscale, stable invariants relevant to the task at hand. This approach is particularly relevant to data generated by physical systems, as such data must respect underlying physical laws. We illustrate this point through many body physics, in which scattering transforms can be loosely interpreted as the machine learning version of fast multipole methods (FMMs). Unlike FMMs, which efficiently simulate the physical system, the scattering transform efficiently learns physical kernels from given states of the system. The resulting learning algorithm obtains state of the art numerical results for the regression of molecular energies in quantum chemistry, obtaining errors on the order of more costly quantum mechanical approaches.

Convergence of gradient descent for kernel minimum error entropy principle

Ting Hu
Wuhan University, China
tinghu@whu.edu.cn

Information theoretical learning (ITL) refers to a framework of learning methods that use concepts of entropies and divergences from information theory to substitute the conventional statistical descriptors of variances and covariance. It becomes an important research topic in signal processing and machine learning as many algorithms have been developed within this framework and many applications domains have been discovered. We study a kernel version of minimum error entropy methods that can be used to find non-linear structures in the data. We show that the kernel minimum error entropy can be implemented by kernel based gradient descent algorithms with or without regularization.

A constructive approach to L0-penalized linear regression

Jian Huang
The Hong Kong Polytechnic University, Hong Kong
j.huang@polyu.edu.hk

We develop a constructive approach to estimating sparse, high-dimensional linear regression models. The approach is a computational algorithm motivated from the KKT conditions for the L_0-penalized least squares solutions. It generates a sequence of solutions iteratively, based on support detection using primal and dual information and root finding. We refer to the algorithm as SDAR for brevity. Under a sparse Rieze condition on the design matrix and certain other conditions, we show that with high probability, the $L2$ estimation error of the solution sequence decays exponentially to the minimax error bound in $O(J\sqrt{\log(R)})$ steps; and under a mutual coherence condition and certain other conditions, the L^{∞} estimation error decays to the optimal error bound in $O(\log(R))$ steps, where J is the number of important predictors, R is the relative magnitude of the nonzero target coefficients. Computational complexity analysis shows that the cost of SDAR is $O(np)$ per iteration. Moreover the oracle least squares estimator can be exactly recovered with high probability at the same cost if we know the sparsity level. We also consider an adaptive version of SDAR to make it more practical in applications. Numerical comparisons with Lasso, MCP and greedy methods demonstrate that SDAR is competitive with or outperforms them in accuracy and efficiency. This is joint work with Yuling Jiao, Yanyan Liu and Xiliang Lv.

Challenges in data science and a possible roadmap of future work

Xiaoming Huo
Georgia Institute of Technology, USA
huo@gatech.edu

The big data is a well-known phenomenon in the modern world. The emerging discipline of data science has inspired a lot of discussion and debate in the scientific research communities, including the mathematical and statistical science community. Contributing to this discussion, in the first part of this talk, I will present a discussion as well as a selective survey on the landscape of data science, as it is forming its foundation. On the second part of this talk, I will present one of my own research, which addresses a particular issue in the enormous spectrum of data science. More specifically, we study how to generate a statistical inference procedure that is both computational efficient and having theoretical guarantee on its statistical performance. We present numerical comparisons with contemporary approaches to demonstrate its advantages.

Spatially distributed sampling and reconstruction of signals on a graph

Yingchun Jiang
Guilin University of Electronic Technology, China
guilinjiang@126.com

A spatially distributed system contains a large amount of agents with limited sensing, data processing, and communication capabilities. First, we introduce a graph structure for a distributed sampling and reconstruction system by coupling agents in a spatially distributed system with innovative positions of signals. Then, we provide a stability criterion for the design of a robust distributed sampling and reconstruction system against supplement, replacement and impairment of agents. Finally, we propose an exponentially convergent distributed algorithm for signal reconstruction, that provides a suboptimal approximation to the original signal.

Tensor network ranks

Lek-Heng Lim
University of Chicago, USA
lekheng@galton.uchicago.edu

A universal problem in science and engineering is to find a function from some input data. The function may be a solution to a PDE and with given boundary/initial data or a target function to be learned from a training set of data. In modern applications, one frequently encounters situations where the function lies in some state space or hypothesis space of prohibitively high dimension — a consequence of, say, very high accuracy solution or massive training set. A common remedy with newfound popularity is to assume that the function has low rank, i.e., may be expressed as a sum of a small number of separable terms. But this is often a bad assumption with weak justification. We will define a vast generalization of the classical notion of rank (i.e., matrix rank, tensor rank, multilinear rank) — given any undirected graph G, there is a G-rank associated with that graph. In particular, the wildly popular tensor network states in physics (e.g., MPS, TTNS, PEPS, MERA) are nothing more than functions of a fixed G-rank for different choices of G. We will discuss various properties of G-ranks. For instance, we will see that a function may have (arbitrarily) high tensor rank and yet (arbitrarily) low G-rank for some G. This is joint work with Ke Ye.

Distributed semi-supervised learning

Shaobo Lin
Wenzhou University, China
sblin1983@gmail.com

Distributed eemi-supervised kernel ridge regression (DSKRR) applies kernel ridge regression (KRR) to each data subset that is distributively stored on multiple servers to produce an output function, and then takes a weighted average of the individual output functions as a final estimator. Our results show that unlabeled data play important roles in enlarging the number of data subsets in DSKRR. Optimal learning rates are established for DSKRR in the framework of learning theory. Numerical experiments including toy simulations and a music-prediction task are employed to demonstrate our theoretical statements and show the power of unlabeled data in distributed learning.

Restoration of atmospheric turbulence-distorted images via RPCA and Quasiconformal maps

Ronald Lok Ming Lui
Chinese University of Hong Kong, Hong Kong
lmlui@math.cuhk.edu.hk

We address the problem of restoring a high-quality image from an observed image sequence strongly distorted by atmospheric turbulence. A novel algorithm is proposed in this paper to reduce geometric distortion as well as space and time-varying blur due to turbulence. By considering an optimization problem, our algorithm first obtain a sharp reference image and a sub-sampled image sequence containing sharp and mildly distorted image frames with respect to the reference image. The sub-sampled image sequence is then stabilized by applying the Robust Principal Component Analysis (RPCA) on the deformation fields between image frames and warping the image frames by a quasiconformal map associated to the low-rank part of the deformation matrix. After image frames are registered to the reference image, the low-rank part of them are deblurred via a blind deconvolution, and the deblurred frames are then fused with the enhanced sparse part. Experiments have been carried out on both synthetic and real turbulence-distorted video. Results demonstrate our method is effective in alleviating distortions and blur, restoring image details and enhancing visual quality.

Balancing principle in supervised learning for a general regularization scheme

Sergei Pereverzyev
Austrian Academy of Sciences, Austria
sergei.pereverzyev@oeaw.ac.at

We discuss the parameter choice in learning algorithms generated by general regularization scheme. In contrast to classical deterministic regularization, the performance of regularized learning algorithms is influenced not only by the smoothness of a target function, but also by the capacity of a regularization space. In supervised learning both the smoothness and the capacity are intrinsically unknown. Therefore, we are interested in a posteriori regularization parameter choice rules and propose a new form of the balancing principle. We provide the analysis of the proposed rule and demonstrate its advantages in simulations.

Joint research with Peter Mathe (WIAS-Berlin) and Shuai Lu (Fudan University, Shanghai).

Adaptive Nyström subsampling for dealing with Big Data

Sergiy Pereverzyev Jr.
University of Innsbruck, Austria
Sergiy.Pereverzyev@uibk.ac.at

In many scientific contexts, one is interested in a reconstruction of the input-output relationship. This relationship can be described by the so-called target function that is in principle unknown. In the statistical learning theory, one considers methods for constructing approximations of the unknown target function using a training data set that contains samples of inputs and corresponding outputs.

In the case of Big Data, training data sets contain a very big number of samples, which makes a direct application of the standard methods almost impossible or very difficult. The Nyström subsampling is one of the techniques that is designed to deal with very big training data sets. This technique can be seen as a regularized projection scheme, and it has two tuning parameters: the so-called regularization parameter, and the dimension of the image of the projection. In the framework of the Nyström subsampling, the dimension of the image of the projection corresponds to the subsampling size.

To the best of our knowledge, currently, in the literature, only a priori choice of the mentioned tuning parameters can be found. This choice requires information about the smoothness of the target function and the capacity of the so-called hypothesis space, which is the space where one looks for the approximation of the unknown target function. This choice can seldom be implemented in practice because the required information is usually not available.

In this talk, we will present an a posteriori strategy for the choice of the tuning parameters in the Nyström subsampling that does not need the mentioned concepts, and nevertheless gives an optimal choice of the tuning parameters.

This work is supported by the Austrian Science Fund (FWF): project P 29514-N32.

Sparse phase retrieval of one-dimensional signals by Prony's method

Gerlind Plonka
University of Goettingen, Germany
plonka@math.uni-goettingen.de

In this talk, we consider the continuous one-dimensional sparse phase retrieval problem to reconstruct a complex-valued signal from the modulus of its Fourier transform. Applications of this problem occur in electron microscopy, wave front sensing, laser optics as well as in X-ray crystallography and speckle imaging. Usually, the challenge in solving one-dimensional phase retrieval problems is to overcome the strong ambiguousness by determining appropriate further information on the solution signal. We show that sparse signals f representable as a linear combination of a finite number N of spikes at arbitrary real locations or as a finite linear combination of B-splines of order m with arbitrary real knots can be almost surely recovered from $\mathcal{O}(N^2)$ intensity measurements up to trivial ambiguities. The constructive proof consists of two steps, where in the first step the Prony method is applied to recover all parameters of the autocorrelation function and in the second step the parameters of f are derived. Moreover, we present an algorithm to evaluate f from its Fourier intensities and illustrate it at different numerical examples.

Adaptive orthonormal systems for matrix-valued functions

Tao Qian
University of Macau, Macau
tqian1958@gmail.com

The talk will be based on a recently established result by DANIEL ALPAY, FABRIZIO COLOMBO, TAO QIAN, AND IRENE SABADINI on analytic approximation to matrix-valued signals based on the so called Adaptive Fourier Decomposition for one-complex-valued signals.

Frames of uniform subframe bounds with applications to erasures

Wenchang Sun
Nankai University, China
sunwch@nankai.edu.cn

We consider a class of frames for which the subframes consisting of finitely many of their elements have common frame bounds depending only on the number of elements in the subframes. We show that such frames are optimal for solving some erasure problems.

Calibration of ε-insensitive loss in support vector machines regression

Hongzhi Tong
University of International Business and Economics, China
tonghz@uibe.edu.cn

We address support vector machines regression (SVMR) based on the ε-insensitive loss. For general distributions and a fixed $\varepsilon > 0$, we show the minimizer of the ε-insensitive loss risk is a set-valued function called conditional ε-median. A calibration inequality is established to describe the relation between the difference of the ε-median function and its estimate in some L^r-space and the corresponding excess risk. This inequality also ensures us to present a nontrivial variance-expectation bound for ε-insensitive loss, and which is known to be important in statistical analysis of the regularized learning algorithms. With the help of the calibration inequality and variance-expectation bound, we finally derive an explicit learning rate of SVMR in L^r-norm under some conventional conditions.

Atomic signal decomposition via SuperEMD

C.K. Chui, H.N. Mhaskar, M.D. van der Walt*
Vanderbilt University, USA
maryke.thom@gmail.com

Decomposition of signals into finitely many primary building blocks, called *atoms*, is a fundamental problem in signal processing, especially when the instantaneous frequencies of the atoms are close together. In this talk, we describe a novel data-driven, local method to address this problem, called *SuperEMD*, which is in essence a clever adaptation and combination of the popular empirical mode decomposition (EMD) and the signal separation operator (SSO). The highlights of our discussion include a natural formulation of the data-driven atoms, a modified sifting process for EMD for real-time implementation with specific reference to handling boundary artifacts, motivation of the SSO, the description of our SuperEMD, and experimental results.

Ensemble boosting algorithms based on data sorting for semi-supervised classification

Jianzhong Wang
Sam Houston State University, Texas, USA
jzwang@shsu.edu

In the presentation, several novel ensemble boosting algorithms based on data sorting for semi-supervised learning (SSL) are introduced. In such an algorithm, we first sort the data into several strings with randomly chosen heads, and construct a pre-classifier as a parallel ensemble of several weak classifiers that are constructed on individual data strings. Then we employ the pre-classifiers in a recursive boosting algorithm to enlarge the labeled data set or to learn the metric on data. We construct the final classifier according to the enlarged labeled set or the learned metric. We display three different boosting algorithms in the talk: Boosting the labeled set at random, boosting the labeled set by clusters, and boosting the labeled set by update metric. The validity and effectiveness of the algorithms are confirmed by the experiments on data sets of different types, such as handwritten digits and hyperspectral images. Comparing to several other popular SSL methods, the proposed method produces very promising results.

Gradient-induced model-free variable selection

Junhui Wang
City University of Hong Kong, Hong Kong
junhwang@cityu.edu.hk

Variable selection is central to sparse modeling, and many methods have been proposed under various model assumptions. In this talk, we will present a model-free variable selection method that allows for general variable effects. As opposed to most existing methods based on an explicit functional relationship, the proposed method attempts to identify non-informative variables that are conditional independent with the response by simultaneously examining the sparsity in multiple conditional quantile functions. It does not require specification of the underlying model for the response, which is appealing in sparse modeling with a relatively large number of variables. The proposed method is implemented via an efficient computing algorithm, which couples the majorize-minimization algorithm and the proximal gradient descent algorithm. The effectiveness of the proposed method is also supported by a variety of simulated and real-life examples. Its asymptotic estimation and variable selection consistencies are established without explicit model assumption.

Kernel-based reconstructions for parametric PDEs

Holger Wendland
University of Bayreuth, Germany
Holger.Wendland@uni-bayreuth.de

In uncertainty quantification, an unknown quantity has to be reconstructed which depends typically on the solution of a partial differential equation. This partial differential equation itself may depend on parameters, some of them are deterministic and some are random. To approximate the unknown quantity one therefore has to solve the partial differential equation (usually numerically) for several instances of the parameters and then reconstruct the quantity from these simulations. As the number of parameters may be large, this becomes a high-dimensional reconstruction problem.

In this talk, I will address the topic of reconstructing such unknown quantities using kernel-based reconstruction methods on sparse grids. I will introduce into the topic, explain the reconstruction process and provide error estimates.

This talk is based upon joint work with C. Rieger (Bonn) and R. Kempf (Bayreuth).

Bias correction for regularized kernel regression with applications in distributed learning

Qiang Wu
Middle Tennessee State University, USA
qwu@mtsu.edu

Regularization kernel network (RKN) is an effective and widely used kernel method for nonlinear regression analysis. We characterize the bias of RKN and propose an approach to correct the bias. The leads to a new kernel method called bias corrected regularization kernel network (BCRKN). BCRKN has smaller asymptotic bias as expected and slightly larger variance as a price of bias reduction. In regression analysis with a single data set, BCRKN achieves the same learning rate as RKN and relaxes the saturation effect if the regularization parameter is appropriately tuned. Empirically they show comparable performance. In distributed regression where the data consist of multiple blocks, BCRKN can reach the optimal learning rate and is empirically more efficient due to bias reduction. (This talk is based on joint works with Zhenchu Guo, Lei Shi, and Hongwei Sun.)

Mathematics in data science

Yuesheng Xu
Zhongshan University, China
yxu06@syr.edu

We shall discuss several crucial mathematical problems that arise in data science, such as recovery information from raw data, machine learning, representation of data and solving non-smooth optimization problems.

The minimal measurement number for low-rank matrix recovery

Zhiqiang Xu
Chinese Academy of Sciences, China
xuzq@lsec.cc.ac.cn

In this talk, we presents several results that address a fundamental question in low-rank matrices recovery: how many measurements are needed to recover low rank matrices? We begin by investigating the complex matrices case and show that $4nr - 4r^2$ generic measurements are both necessary and sufficient for the recovery of rank-r matrices in $\mathbb{C}^{n\times n}$. Thus, we confirm a conjecture which is raised by Eldar, Needell and Plan for the complex case. We next consider the real case and prove that the bound $4nr - 4r^2$ is tight provided $n = 2^k + r, k \in \mathbb{Z}_+$. Motivated by Vinzant's work, we construct 11 matrices in $\mathbb{R}^{4\times 4}$ by computer random search and prove they define injective measurements on rank-1 matrices in $\mathbb{R}^{4\times 4}$. This disproves the conjecture raised by Eldar, Needell and Plan for the real case. We also apply this method to generalized phase retrieval.

Adaptive False Discovery Rate regression with application in integrative analysis of large-scale genomic data

Can Yang
Hong Kong Baptist University, Hong Kong
eeyang@hkbu.edu.hk

To address scientific questions, we often design experiments and collect data from experiments. Conventionally, we often focus on the data set at hand and improve analysis results by refining models. The rising of Big Data may change the way of doing research What if combining our data at hand with other existing information that hides in the Big Data Mountain?

In this talk, we consider a large-scale testing problem in genomic data analysis. Recent international projects, such as the Encyclopedia of DNA Elements (ENCODE) project, the Roadmap project and the Genotype-Tissue Expression (GTEx) project, have generated vast amounts of genomic annotation data, e.g., epigenome and transcriptome. There is great demanding of effective statistical approaches to integrate genomic annotations with the results from genome-wide association studies (GWAS). To explore genetic architecture of human complex phenotypes, rather than only relying on GWAS, we introduce Adaptive False Discovery Rate (AdaFDR) regression to integrate genomic annotations with GWAS. For a given phenotype, not only AdaFDR increase the power of mapping its risk variants, but also adaptively incorporates relevant annotations for prioritization of genetic risk variants, allowing nonlinear effects among these annotations, such as interaction effects between genomic features. The developed algorithm is scalable to genome-wide analysis. Using AdaFDR, we performed integrative analysis of genome-wide association studies on human complex phenotypes and genome-wide annotation resources, e.g., Roadmap epigenome. The analysis results revealed interesting regulatory patterns of risk variants, offering new biological insights on genetic architectures of complex phenotypes.

Recursive schemes for shape functions in the mode decomposition problem

Haizhao Yang
Duke University, USA
haizhao@math.duke.edu

This talk is about a novel recursive diffeomorphism-based scheme for one-dimensional generalized mode decomposition problem that aims at extracting generalized modes $\alpha_k(t)$ $\times s_k(2\pi N_k \phi_k(t))$ from their superposition $\sum_{k=1}^{K} \alpha_k(t) s_k(2\pi N_k \phi_k(t))$, assuming that amplitude and phase functions $\alpha_k(t)$ and $N_k \phi_k(t)$ are known. We introduce a few newly developed approaches to estimate wave shape functions $s_k(t)$ and prove the convergence of the recursive scheme. If time permitted, a fast algorithm will be introduce to speed-up the calculation and convergence.

Questions on mono-components

Lihua Yang
Sun Yat-sen University, China
mcsylh@mail.sysu.edu.cn

Mono-component is a widely used concept in non-stationary signal processing and time-frequency analysis. However, there is no accepted definition and there are some arguments on this concept. This talk will give a brief introduction and recent developments to this issue. Open questions will also be discussed.

Boosting with structural sparsity – A differential inclusion approach

Yuan Yao
Hong Kong University of Science and Technology, Hong Kong
yuany@ust.hk

Boosting, as gradient descent method, is arguably the 'best off-the-shelf' methods in machine learning. Here a novel Boosting-type algorithm is proposed based on restricted gradient descent whose underlying dynamics are governed by differential inclusions. In particular, we present an iterative regularization path with structural sparsity where the parameter is sparse under some linear transforms, based on variable splitting and the Linearized Bregman Iteration. Hence it is called Split LBI. Despite its simplicity, Split LBI outperforms the popular generalized Lasso in both theory and experiments. A theory of path consistency is presented that equipped with a proper early stopping, Split LBI may achieve model selection consistency under a family of Irrepresentable Conditions which can be weaker than the necessary and sufficient condition for generalized Lasso. Furthermore, some l2-error bounds are also given at the minimax optimal rates. The utility and benefit of the algorithm are illustrated by applications on image reconstruction, partial order ranking, and Alzheimers disease detection via neuroimaging.

How to implement ADMM to large-scale datasets?

Xiaoming Yuan
Hong Kong Baptist University, Hong Kong
xmyuan@hkbu.edu.hk

The alternating direction method of multipliers (ADMM) is being popularly used for a wide range of applications including many in data science. To tackle very large-scale datasets of some representative statistical learning problems such as the distributed LASSO, it is neither possible nor necessary to solve the ADMM subproblems exactly or up to a high precision — inexact solutions in low-accuracy are cheaper while indeed better towards the convergence. We try to study this implementation issue mathematically. We are particularly interested in the case where the subproblems are very large-scale systems of linear equations and they are solved iteratively by benchmark numerical linear algebra solvers such as CG or SOR. We show how to rigorously ensure the convergence for this case. More specifically, we estimate precisely how many iterations of these numerical linear algebra solvers are needed to ensure the convergence. We thus make the implementation of ADMM embedded with benchmark numerical linear algebra solvers for the large-scale cases of some statistical learning problems fully automatic, with proved convergence.

Multiscale data analysis: Framelets, manifolds and graphs

Xiaosheng Zhuang
City University of Hong Kong, Hong Kong
xzhuang7@cityu.edu.hk

While Big Data are high-volume, high-dimensional, and high complexity, they are typically concentrated on low-dimensional manifolds or can be represented by graphs, digraphs, etc. Sparsity is the key to the successful analysis of data in various forms. Multiscale representation systems provide efficient and sparse representation of various data sets. In this talk, we will discuss the characterizations, construction, and applications of framelets on manifolds and graphs. We shall demonstrate that tight framelets can be constructed on compact Riemannian manifolds or graphs, and fast algorithmic realizations exist for framelet transforms on manifolds and graphs. Explicit construction of tight framelets on the sphere as well as numerical examples will be shown.