Conference Guide

Mini Special Semester on Inverse Problems 2010

„Inverse Problems in Data Driven Modelling"

Organized by
Shuai Lu
Sergiy Pereverzyev Jr.
Lorenzo Rosasco
Sivananthan Sampath

RICAM
JOHANN · RADON · INSTITUTE
FOR COMPUTATIONAL AND APPLIED MATHEMATICS

ÖAW
Austrian Academy of Sciences

Content
Time Table
Abstracts
List of Participants

July 20th - July 22nd, 2010 - RICAM, Linz, Austria

About the Workshop

In many problems in industry and applied sciences, technological progresses made available small, as well as massive, samples of complex high dimensional data that need to be analyzed and summarized in order to extract relevant information. In this context, learning theory and machine learning provided a suitable framework and effective algorithmic solutions for a variety of problems in computer vision, biomedical informatics, natural language processing and computational neuroscience. Most of these problems are inherently ill-posed. On the other hand, the regularization theory gives powerful tools for solving general inverse and ill-posed problems. In this workshop, we aim at bringing together researchers working on learning, regularization and data modeling problems to present the state of the art in each field, discuss future challenges, and foster fruitful collaborations.

Organizers

Shuai Lu, Johann Radon Institute, Austria
Sergiy Pereverzyev Jr. (Chair), Johannes Kepler University Linz, Austria
Lorenzo Rosasco, Massachusetts Institute of Technology, USA
Sivananthan Sampath, Johann Radon Institute, Austria

Venue

Johann Radon Institute, Hochschulfondsgebäude, Room: HF 9901, July 20-July 22, 2010.

Invited Speakers

Andreas Argyriou, Toyota Technological Institute at Chicago, USA

Gilles Blanchard, Weierstraß Institut (WIAS), Germany

Christine De Mol, Universite Libre de Bruxelles, Belgium

Alexander Goldenshluger, University of Haifa, Israel

Gerard Kerkyacharian, Université Paris VII, France

Vera Kurkova, Institute of Computer Science, Czech Republic

Jean-Michel Loubes, Université de Toulouse 3, France

Boaz Nadler, Weizmann Institute of Science, Israel

Guillaume Obozinski, INRIA - WILLOW Project-team, France

Gianluigi Pillonetto, University of Padova, Italy

Massimiliano Pontil, University College London, UK

Franciszek Rakowski, University of Warsaw, Poland

Katerina Schindler, Austrian Academy of Sciences, Austria

Milan Stehlik, University of Linz, Austria

Gabriele Steidl, University of Mannheim, Germany

Alessandro Verri, Università degli Studi di Genova, Italy

Time Schedule

Time	Tuesday 20 July
9.00 – 9.30	Opening
9.30 – 10.20	De Mol
10.20 – 10.50	Break
10.50 – 11.40	Goldenshluger
11.40 – 14.00	Lunch
14.00 – 14.50	Verri
14.50 – 15.40	Blanchard
15.40 – 16.10	Break
16.10 – 17.00	Stehlik
19.00	Conference Dinner

Time	Wednesday 21 July	Thursday 22 July
9.00 – 9.50	Kerkyacharian	Kurkova
9.50 – 10.40	Steidl	Argyriou
10.40 – 11.10	Break	
11.10 – 12.00	Rakowski	Pilonetto
12.00 – 14.00	Lunch	
14.00 – 14.50	Nadler	Loubes
14.50 – 15.40	Obozinski	Schindler
15.40 – 16.10	Break	
16.10 – 17.00	Rosasco	Pereverzyev Jr.

Abstracts

Spectral Regularization in Machine Learning

ANDREAS ARGYRIOU
Toyota Technological Institute at Chicago, USA

Multi-task learning and transfer learning extend the standard paradigm of supervised learning. In multi-task learning, samples for multiple related tasks are given and the goal is to learn a function for each task while generalizing well (transferring learned knowledge) on new tasks. I will present some formulations of such problems as convex programs with matrix variables. Of special interest are some recent approaches based on regularization with Schatten Lp norms, such as the trace (nuclear) norm. I will also discuss algorithms and theoretical bounds for these problems.

On optimal rates for kernel conjugate gradient regularization under random design and noise

GILLES BLANCHARD
Weierstrass Institute for Applied Analysis and Stochastics, Berlin, Germany

Conjugate gradient (CG) with early stopping in a least-squares fitting context is a form of regularization which is used in practice in the so-called Partial Least Squares algorithm. We study convergence rates of this type of regularization when applied in a reproducing Hilbert kernel space. One important difference with other approaches (regularized risk minimization or linear regularization operators) is that even for a fixed number of steps k, the obtained estimator is nonlinear.

We prove optimal (up to log factor) rates of convergence in the statistical sense under source conditions (assumed to be known) on the true regression function, and when this function belongs the reproducing kernel Hilbert space. If the latter assumption is not fulfilled, we also obtain comparable convergence rates if additional unlabeled data are available. The rates in these two cases match those obtained by A.Caponnetto for linear regularization operators in the same "learning theoretic" setting.

The results rely on a reworking of arguments initially developed by A. Nemirovskii in his classical work on CG regularization for (deterministic) inverse problems.

Catching features with a lasso or an elastic net?

CHRISTINE DE MOL
Université Libre de Bruxelles, Belgium

I will describe feature selection methods based on penalized regression, and in particular the "lasso" (L^1 penalty) and the "elastic net" (combined L^1 and L^2 penalty). I will summarize the theoretical background in the framework of learning theory and enlighten connections with regularization theory for inverse problems. To illustrate the concepts, I will focus on two specific applications: face detection in computer vision and gene selection in bioinformatics.

Adaptive nonparametric estimation by selection of estimators

ALEXANDER GOLDENSHLUGER
University of Haifa, Israel

The talk is about adaptive nonparametric estimation of multivariate functions from noisy observations. The problem of adaptive estimation is formulated as that of selection from a family of linear estimators. We propose a general selection rule and derive oracle inequalities. It is shown that the rule leads to adaptive minimax estimators in a wide variety of estimation setting. In particular, the resulting estimators adapt both to the smoothness and unknown structure of the function to be estimated. The talk is based on joint works with O. Lepski.

Some recent bounds on minimal singular value: a comparison

KATEŘINA HLAVÁČKOVÁ-SCHINDLER
Commission for Scientific Visualization, Austrian Academy of Sciences, Austria

We present a new lower bound on minimal singular values of real matrices based on Frobenius norm and determinant. We show that under certain assumptions on the matrix is our estimate sharper than three recent upper bounds for minimal singular values based on a matrix norm and determinant, namely the upper bounds from Hong and Pan, from Piazza and Politi and from Hou-Biao Li et al.

Radon needlet thresholding

GERARD KERKYACHARIAN
Université Paris VII, France

We provide a new algorithm for the treatment of the noisy inversion of the radon transform using an appropriate thresholding technique in a well chosen new basis. We establish a minimax result and prove its optimality.

Integral transforms induced by computational units

VĚRA KŮRKOVÁ
Institute of Computer Science, Academy of Sciences of the Czech Republic

Integral transformations play an important role in applied science. Often such transforms have the forms of integral operators with various kernels. Also functions computable by various computational units can be used as kernels of integral operators. Metaphorically, such integrals can be interpreted as computational models with infinitely many units. We show that such infinite models can be used as useful tools in mathematical theory of neurocomputing and learning. For widely used types of computational units (such as sigmoidal perceptrons or Gaussian kernel units), large classes of multivariable functions can be represented as infinite networks with these units. We derive estimates of rates of approximation of such functions by finite networks with increasing number of units. We use these estimates to obtain upper bounds on speed of convergence of suboptimal infima of errors functionals over networks with increasing number of units to global minima of such functionals, which are described by methods from theory of inverse problems. We also obtain some estimates of dependence of approximation errors on input dimension and derive some comparisons of capabilities of units of various types. This work was supported by CA CR grant 201/08/1744.

Tests for inverse problems: a direct or indirect problem?

JEAN-MICHEL LOUBES
Université de Toulouse 3, France

One of the main difference between direct and indirect problems comes from the fact that two spaces are at hand: the space of the observations and the space where the function will be estimated, namely, the operator mapping one space into another. In this talk, we try to provide an answer to the natural question whether tests for inverse problems require a particular attention, in the sense that they should be built taking into account the operator of inverse problem, or whether direct tests on the observations, i.e. without

inverting the operator, could be used instead. Hence we focus on inference on rough data without inverting the operator or using regularization methods.

Semi-supervised Learning, global vs. multiscale methods and harmonic analysis

BOAZ NADLER
Weizmann Institute of Science, Rehovot, Israel

In recent years there is increasing interest in learning from both labeled and unlabeled data (a.k.a. semi-supervised learning, or SSL). The key assumption in SSL, under which an abundance of unlabeled data may help, is that there is some relation between the unknown response function to be learned and the marginal density of the predictor variables. In the first part of this talk I'll present a statistical analysis of two popular graph based SSL algorithms: Laplacian regularization method and Laplacian eigenmaps. In the second part I'll present a novel multiscale approach for SSL as well as supporting theory. Some intimate connections to harmonic analysis on abstract data sets will be discussed.

Joint work with Nati Srebro (TTI), Xueyuan Zhou (Chicago), and with Matan Gavish (WIS/Stanford) Ronald Coifman (Yale).

Support union recovery in high-dimensional multivariate linear regression

GUILLAUME OBOZINSKI
INRIA - WILLOW Project-team, France

Regularization of the L^1 norm provides a computationally efficient approach to variable selection, which motivated a large body of work in the recent years. In particular, L^1 regularization achieves with high probability the correct recovery of the set of variables forming the support of a linear regression, under conditions that are today reasonably well understood. In this talk, I will consider the more general problem of variable selection in multivariate regression, where the parameters form a matrix B* of size p by K corresponding to input and output dimensions. In that context, it is reasonable to assume that many variables are irrelevant to all of the K output predictions so that only a few rows of B* are not identically zero. I will focus on the estimation, in high-dimension, of the set of non-zero rows. Considering an extension of L^1 - regularization that penalizes the sum of l^2 norms of the rows of B*, it is possible to show that the success of the penalization depends on a function of the true parameters and the design called the "sparsity-overlap" function. Based on that result I will discuss under which conditions the regularization considered performs better than the L^1 - regularization and by how much.

Dual regularized total least squares for the prediction problem in learning theory

SERGIY PEREVERZYEV JR.
Johannes Kepler University of Linz, Austria

The prediction problem in Learning Theory consists in predicting the future outputs of a system using the system outputs observed in the past. We consider a system for which the time evolution of its output can be described by a real-valued function of one variable. For example in the human body, the blood glucose concentration is such an output. In this case, in the prediction problem one would like to estimate the function that models the time evolution of the system output using the available observations of this output. The detailed formulation of the prediction problem involves an operator equation for the unknown function whose numerical solution is unstable.

Our analysis of the nature of this instability shows that the involved operator equation has a noisy operator and a noisy right hand side. The Dual Regularized Total Least Squares (DRTLS), a recently developed method, is designed for such equations. Based on our analysis, we think that the DRTLS is an adequate tool for solving the prediction problem.

We present the general application of the DRTLS to the prediction problem, discuss its several particular realizations, and present results of the numerical experiments with this method.

The talk presents the results of the joint work with Shuai Lu and Sivananthan Sampath (RICAM).

New nonparametric approaches to linear system identification

GIANLUIGI PILLONETTO
University of Padova, Italy

The classical approach to system identification postulates a set of parametric models and the most suitable one is chosen by complexity criteria such as Akaike (AIC) or Bayesian Information Criterion (BIC). In this talk I will discuss alternative nonparametric paradigms for linear system identification which exploit the framework of regression via Gaussian processes. Instead of postulating finite-dimensional models, the input-output relationship is searched in suitable infinite-dimensional reproducing kernel hilbert spaces which encode information on system and/or predictor stability and will be fully characterized. Numerical examples will illustrate advantages of the new approaches over the classical parametric ones, in particular in terms of robustness relative to model order selection.

Multi-task Learning: Theory and Practice

MASSIMILIANO PONTIL
University College London, UK

We discuss the problem of estimating a structured matrix with a large number of elements. A key motivation for this problem occurs in multi-task learning. In this case, the columns of the matrix correspond to the parameters of different regression or classification tasks, and there is structure due to relations between the tasks. We present a general method to learn the tasks' parameters as well as their structure. Our approach is based on solving a convex optimization problem, involving a data term and a penalty term. We highlight different types of penalty terms which are of practical and theoretical importance. They implement structural relations between the tasks and achieve a sparse representations of parameters. We address computational issues as well as the predictive performance of the method. Finally we discuss how these ideas can be extended to learn non-linear task functions by means of reproducing kernels.

Agent Based Modelling in studying the epidemic spread processes on the country-wide scale

FRANCISZEK RAKOWSKI
Interdiscyplinary Centre for Mathematical and Computational Modelling,
University of Warsaw, Poland

A construction of an agent-based model for studying the effects of air-borne diseases epidemic spread will be described. Resulting various scenarios emerging from large scale stochastic simulations of disease spread in Poland will be discussed. Simple transportation rules were employed to mimic individuals' travels in dynamic route-changing schemes, allowing for the infection spread during a journey. The work was based on national Census Burreau data regarding the social contact structure, and the LandScan data for population density. Various techniques used for de-aggregation of the data, to combine resulting bi-partite graphs into social contact network will be presented. Parameter space was checked for stable behaviour, especially towards the effective infection transmission rate's variability. Application of the model allowed to observe two different types of epidemic scenarios: characteristic for urban and rural areas. Based on this some important factors, e.g. attack rate, burden of disease, etc. for regional public health authorities can be determined.

Inverse Problems in Cancer Modelling

MILAN STEHLÍK
Johannes Kepler University Linz, Austria

When we consider fractal based cancer diagnostic, many times a statistical procedure to assess the fractal dimension is needed. We shall look for some analytical tools for discrimination between cancer and healthy ranges of fractal dimensions of tissues. [1] discussed planar tissue preparations in mice which has a remarkably consistent scaling exponents (fractal dimensions) for tumor vasculature even among tumor lines that have quite different vascular densities and growth characteristics. In our talk we will present two important inverse problems:

1) We may construct a continuously parameterized set of weakly (sometimes strongly) consistent estimators of Pareto tail (and consequently cancer risk). The one alternative to the Hill estimator will be introduced and discussed. Therefore we may get a relatively large scale of possible estimated risk levels, which constitutes an inverse problem.

2) Scaling of the data. This is apparently important for the power and p-value modelling. Scaling may led to a scale of p-values and powers, which constitutes an inverse problem.

Beside 1) and 2) non invasive techniques generally may produce inverse problems, e.g. estimating a Hausdorff fractal dimension from boundary of examined tissue. We will discuss these issues in the context of our recent results (see e.g. [2]).

REFERENCES

[1] Baish J.W. and Jain R.K. (2009) Fractals and cancer. Cancer Research, 60, 3683-3688.

[2] Filus J., Filus L. and Stehlík M. (2009). Pseudoexponential modelling of cancer diagnostic testing, Biometrie und Medizinische Informatik, Greifswalder Seminarberichte Heft 15, 41-54.

Operator Splitting Methods in Image Processing

GABRIELE STEIDL
University of Mannheim, Germany

Operator splitting techniques were developed in the 60th for the efficient solution of linear systems of equations appearing in connection with the numerical solution of partial differential equations. More than 20 years later these methods were generalized to nonlinear, set-valued, monotone operators and were recently successfully applied in image processing.

After motivating why methods of convex analysis/non-smooth convex optimization are useful for many image restoration tasks we give a brief introduction into the relevant operator splitting techniques. We point out that for special functionals there exist equivalent derivations via averaged operators, (Bregman) proximal point methods or augmented Lagrangian methods.

Then we present applications of operator splitting methods in direction-steered inpainting, denoising of images corrupted with non-additive noise and dithering.

Spectral Methods for Learning High Dimensional Data

LORENZO ROSASCO
Massachusetts Institute of Technology, USA

In this talk we discuss our recent advances to develop machine learning techniques for the analysis of high dimensional data sets arising in a wide variety of applications.

The tools we build upon are spectral and analytical methods and allow to model complex data, while achieving good computational and theoretical properties. Empirically the derived algorithms obtain state of the art performances both on simulated and real data.

Interestingly, our approach highlights the connections between learning theory and other fields in applied sciences, such as inverse problems, statistics and signal processing.

Learning, Regularization and Ill-posed Inverse Problems

ALESSANDRO VERRI
Università degli Studi di Genova, Italy

In this talk we will start reviewing some of our results establishing a connection between learning theory and the theory of ill-posed inverse problems and then discuss some more recents advances.

List of Participants

Andreas Argyriou
Toyota Technological Institute at Chicago, USA
argyriou@ttic.edu

Sven Beuchler
RICAM – Austrian Academy of Sciences, Austria
sven.beuchler@oeaw.ac.at

Gilles Blanchard
Weierstraß Institut (WIAS)
Germany
gilles.blanchard@wias-berlin.de

Christine De Mol
Universite Libre de Bruxelles
Belgium
demol@ulb.ac.be

Alexander Goldenshluger
University of Haifa
Israel
goldensh@stat.haifa.ac.il

Gerard Kerkyacharian
Université Paris VII
France
kerk@math.jussieu.fr

Vera Kurkova
Institute of Computer Science
Czech Republic
vera@cs.cas.cz

Jean-Michel Loubes
Université de Toulouse 3
France
loubes@math.univ-toulouse.fr

Shuai Lu
RICAM – Austrian Academy of Sciences, Austria
shuai.lu@oeaw.ac.at

Boaz Nadler
Weizmann Institute of Science
Israel
boaz.nadler@weizmann.ac.il

Valeriya Naumova
RICAM – Austrian Academy of Sciences, Austria
valeriya.naumova@oeaw.ac.at

Trung Thanh Nguyen
RICAM – Austrian Academy of Sciences, Austria
trung-thanh.nguyen@oeaw.ac.at

Guillaume Obozinski
INRIA, Laboratoire d'Informatique de l'Ecole Normale Supérieure, France
guillaume.obozinski@ens.fr

Sergiy Pereverzyev Jr.
Johannes Kepler Universität
Austria
pereverzyev@indmath.uni-linz.ac.at

Gianluigi Pillonetto
University of Padova
Italy
giapi@dei.unipd.it

Massimiliano Pontil
University College London
UK
m.pontil@cs.ucl.ac.uk

Franciszek Rakowski
University of Warsaw
Poland
F.Rakowski@icm.edu.pl

Lorenzo Rosasco
Massachusetts Institute of Technology
USA
lrosasco@mit.edu

Sivananthan Sampath
RICAM – Austrian Academy of
Sciences, Austria
sivananthan.sampath@oeaw.ac.at

Katerina Schindler
Austrian Academy of Sciences
Austria
katerina.schindler@assoc.oeaw.ac.at

Milan Stehlik
University of Linz
Austria
Milan.Stehlik@jku.at

Gabriele Steidl
University of Mannheim
Germany
steidl@math.uni-mannheim.de

Alessandro Verri
Università degli Studi di Genova
Italy
verri@disi.unige.it

Mariya Zhariy
Johannes Kepler Universität
Austria
zhariy@indmath.uni-linz.ac.at