

MATHEMATISCHES FORSCHUNGSINSTITUT OBERWOLFACH

Report No. 31/2012

DOI: 10.4171/OWR/2012/31

Learning Theory and Approximation

Organised by
Kurt Jetter, Hohenheim
Steve Smale, Hong Kong
Ding-Xuan Zhou, Hong Kong

June 24th – June 30th, 2012

ABSTRACT. Learning theory studies data structures from samples and aims at understanding unknown function relations behind them. This leads to interesting theoretical problems which can be often attacked with methods from Approximation Theory. This workshop - the second one of this type at the MFO - has concentrated on the following recent topics: Learning of manifolds and the geometry of data; sparsity and dimension reduction; error analysis and algorithmic aspects, including kernel based methods for regression and classification; application of multiscale aspects and of refinement algorithms to learning.

Mathematics Subject Classification (2000): 68Q32, 41A35, 41A63, 62Jxx.

Introduction by the Organisers

The workshop *Learning Theory and Approximation*, organised by Kurt Jetter (Stuttgart-Hohenheim), Steve Smale (Hong Kong) and Ding-Xuan Zhou (Hong Kong) was held June 24–30, 2012. The meeting was well attended with 47 participants from Asia, Europe and North America. It provided an excellent platform for fruitful interactions among scientists from learning theory and approximation theory.

The first part of the scientific program consisted of a few talks on learning geometric structures from data. Steve Smale's talk on mathematical foundations of immunology demonstrated strong connections on data analysis for peptides and amino acid chains among the research fields of computational biology and geometry, learning theory and approximation theory. Nat Smale presented a Hodge

theory for Alexandrov spaces with curvature bounded above including Riemannian manifolds, Riemannian manifolds with boundary and singular spaces such as simplicial complexes whose faces have constant curvature, and Tits buildings. The described Hodge decomposition can be applied to data analysis and processing. Modelling data by manifolds reflects many important aspects of realistic data and provides a direct connection with differential geometry. In particular, manifold learning algorithms based on graph Laplacians constructed from data have received considerable attention both in practical applications and theoretical analysis. Belkin discussed the behavior of graph Laplacians at points at or near boundaries, intersections and edges, and showed that the behavior of graph Laplacians near these singularities is quite different from that in the interior of the manifolds. Jost spoke about the topic of geometric structures on the space of probability measures in the area of information geometry, and introduced a sufficient statistic for a parametrized family of measures under which the Fisher metric and the Amari-Chentsov tensors remain invariant. Von Luxburg talked about the problem of density estimation from unweighted k -nearest neighbor graphs, and connections to graph learning algorithms like spectral clustering or semi-supervised learning. Lim gave a talk on principal components of cumulants, discussing the geometry underlying cumulants and examining two ways to their principal components analysis, decomposing a homogeneous form into a linear combination of powers of linear forms, and decomposing a symmetric tensor into a multilinear combination of points on a Stiefel manifold.

Sparsity is an important property for dimension reduction, data representation and analysis, and information retrieval. In this workshop, some statisticians and approximation theorists discussed sparsity for various purposes and raised interesting problems for approximation theory: Tsybakov introduced a compound functional model as a nonparametric generalization of the high-dimensional linear regression model under the sparsity scenario and presented minimax rates of convergence in terms of structural conditions of functions. Dahmen applied deep analysis from tree-structured approximation to classification algorithms with adaptive partitioning and analyzed their risk performance. The analysis allows one to relax classical Hölder smoothness to weaker Besov smoothness, which leads to interesting approximation theory problems. Sparsity was a core issue in classical support vector machines. Christmann's talk was focussed on the question how to draw statistical decisions based on nonparametric methods such as bootstrap approximations of support vector machines and qualitatively robust support vector machines. His discussions on various loss functions raised research problems for approximation theorists. Li considered the compressed sensing topic of nonuniform support recovery via orthogonal matching pursuit from noisy random measurements. Zhou talked about error analysis and sparsity for support vector regression, coefficient-based regularization with ℓ^1 -penalty, and kernel projection machines with ℓ^q -penalty.

Both approximation theory and learning theory provide useful tools for data analysis and statistics. This is reflected by quite a few talks in this workshop.

Wu gave a learning theory perspective on the empirical minimum error entropy (MEE) principle developed in the fields of signal processing and data mining, and he provided a rigorous consistency analysis of some MEE learning algorithms in terms of approximation theory conditions on the model and hypothesis spaces. Döring considered a regression model with a change point in the regression function and investigated the consistency with the increasing sample size of the least squares estimates of the change point, for which the convergence rates depend on the order of smoothness of the regression function at the change point. Minh proposed a regularized spectral algorithm for hidden Markov models with numerical stability, and gave some theoretical justification and simulations on real data from pattern recognition. Pereverzyev applied the least squares Tikhonov regularization schemes in reproducing kernel Hilbert spaces to the practical problem of blood glucose reading, discussed intensively how to choose the hypothesis space, and described a kernel adaptive regularized algorithm.

Kernels have been an essential part of both learning theory and approximation theory. They form the topic of a few talks in this workshop. Plonka introduced Prony's method for solving inverse problems to the workshop audience, and described her recent work on function reconstruction in terms of sparse Legendre expansions and a new perception of Prony's method based on eigenfunctions of linear operators. Steinwart surveyed approximation theory properties of reproducing kernel Hilbert spaces (eigenvalues, entropy numbers, interpolation spaces, Mercer representations) and some related kernel methods for both supervised and unsupervised learning. Schaback described some methods for explicit constructions of new positive definite radial kernels, in particular kernels that are linked to generalized Sobolev spaces. Zu Castell demonstrated some kernel-based methods for learning and approximation. For conditionally positive definite kernels, he raised some approximation theory questions about the associated reproducing kernel Pontryagin space. Rosasco described the problem of learning the region where a probability measure is concentrated by means of separating kernels. Some approximation theory questions are mentioned such as the approximation of sets under the Hausdorff distance and the existence of completely separating kernels.

Approximation theory and ideas of multiscale analysis from wavelets have been applied in learning theory and have further potential applications. The workshop contains quite a few talks discussing these areas and other possible connections between learning and approximation. Kutyniok gave a survey on shearlets and demonstrated their applications in sparse approximation and dictionary learning. Mhaskar talked about function approximation on data dependent manifolds. The research area of irregular sampling was described by Stöckler in his talk. Han's talk was on linear-phase moments in wavelet analysis and approximation theory. Bernstein polynomials and Bernstein-Durrmeyer operators associated with general probability measures together with their applications to learning theory were discussed by Wu and Berdysheva in their talks. The ideas of tracking multiscale structures by subdivision schemes and refinement algorithms together with potential applications in learning theory were discussed by Ebner and Jetter.

The organizers acknowledge the friendly atmosphere provided by the Oberwolfach institute, and express their thanks to the entire staff.

Workshop: Learning Theory and Approximation

Table of Contents

Steve Smale	
<i>Mathematical foundation of immunology</i>	1901
Nat Smale	
<i>A Hodge theory for Alexandrov spaces with curvature bounded above</i> ...	1902
Mikhail Belkin (joint with Qichao Que, Yusu Wang, Xueyuan Zhou)	
<i>Dealing with singular manifolds: theory and applications</i>	1904
Jürgen Jost (joint with Nihat Ay, Hông Vân Lê and Lorenz Schwachhöfer)	
<i>Information geometry and sufficient statistics</i>	1905
Qiang Wu (joint with Jun Fan, Ting Hu, and Ding-Xuan Zhou)	
<i>A Learning Theory perspective on the empirical minimum error entropy principle</i>	1905
Maik Döring (joint with Uwe Jensen)	
<i>Change point estimation in regression models</i>	1907
Gitta Kutyniok (joint with Jakob Lemvig, Wang-Q Lim)	
<i>Shearlets: sparse approximation and dictionary learning</i>	1909
Gerlind Plonka (joint with Thomas Peter)	
<i>Approximation by k-sparse sums of eigenfunctions of linear operators</i> ..	1910
Lek-Heng Lim (joint with Jason Morton)	
<i>Principal components of cumulants</i>	1912
Hà Quang Minh (joint with Marco Cristani, Alessandro Perina, Vittorio Murino)	
<i>A regularized spectral algorithm for Hidden Markov Models</i>	1913
Ulrike von Luxburg	
<i>Can we estimate the density from an unweighted, random k-nearest neighbor graph?</i>	1914
Hrushikesh Mhaskar	
<i>Function approximation on data defined manifolds</i>	1915
Alexandre B. Tsybakov (joint with Arnak S. Dalalyan, Yuri Ingster)	
<i>Statistical inference in compound functional models</i>	1915
Andreas Christmann (joint with Matías Salibíán-Barrera, Stefan Van Aelst, Robert Hable)	
<i>On approximations of the finite sample distribution of Support Vector Machines</i>	1918

Song Li	
<i>Nonuniform support recovery from noisy random measurements by orthogonal matching pursuit</i>	1920
Sergei V. Pereverzyev (joint with V. Naumova, S. Sivananthan)	
<i>Learning in variable RKHSs with application to the blood glucose reading</i>	1920
Ingo Steinwart	
<i>Eigenvalues, entropy numbers, and Mercer representations for RKHSs</i> .	1922
Robert Schaback	
<i>Methods of kernel construction</i>	1924
Wolfgang zu Castell (joint with Georg Berschneider)	
<i>Kernel based methods in Learning and Approximation</i>	1928
Wolfgang Dahmen (joint with Peter Binev, Albert Cohen, Ronald DeVore)	
<i>Classification algorithms using adaptive partitioning</i>	1929
Lorenzo Rosasco (joint with Ernesto De Vito, Alessandro Toigo)	
<i>Learning sets with separating kernels</i>	1931
Joachim Stöckler (joint with Karlheinz Gröchenig)	
<i>Irregular sampling in subspaces of $L_2(\mathbb{R})$</i>	1933
Bin Han	
<i>Linear-phase moments in wavelet analysis and approximation theory</i> ...	1934
Zongmin Wu (joint with Xingping Sun, Limin Ma)	
<i>Sampling scattered data with Bernstein polynomials: stochastic and deterministic error estimates</i>	1936
Elena E. Berdysheva	
<i>Bernstein-Durrmeyer operators with arbitrary weight functions</i>	1938
Ding-Xuan Zhou	
<i>Error analysis and sparsity of some learning algorithms</i>	1940
Oliver Ebner	
<i>Stochastic aspects of nonlinear refinement algorithms</i>	1943
Kurt Jetter (joint with Xianjun Li)	
<i>Non-negative subdivision and Markov chains</i>	1944