HIM Trimester Program on Analysis and Numerics for High Dimensional Problems

Book of Abstracts

Workshop on Theoretical Aspects of High-Dimensional Problems and Information-Based Complexity

Bonn, 20 – 24 June

Organizers: Markus Hegland, Vladimir Pestov, Ian Sloan and Henryk Woźniakowski
In recent years advances have been made to better understand problems in high dimensions from the view of geometric analysis, functional analysis, stochastics and complexity theory. Here, the concentration of measure phenomenon, ultrametric spaces, dimension embedding and reduction, learning theory or information based complexity are well studied, but these different areas developed independently without much interaction. The workshop aims to bring together researchers interested in the areas of geometric and functional analysis, information based complexity and multiscale analysis.

Thursday, June 23, is an official holiday and there will be no talks but visitors are welcome to meet at HIM for collaboration and discussions. Please contact us on the Monday if you are interested in a group excursion on Thursday.

On Friday, June 24, we have talks, exploration of new ideas, tutorials etc which will be organised during the week. Please contact the organisers before Thursday or just advertise on the white-board at the HIM if you wish to present or are interested in a topic to be covered.

Each talk is allocated a one hour slot. Nevertheless, we encourage talks of between 30 and 50 minutes duration, leaving ample time for discussion.

**Workshop Venue**

The workshop is hosted by the Hausdorff Research Institute for Mathematics, Universität Bonn, [http://www.hausdorff-research-institute.uni-bonn.de](http://www.hausdorff-research-institute.uni-bonn.de) in cooperation with the Institute for Numerical Simulation [http://www.ins.uni-bonn.de](http://www.ins.uni-bonn.de).
Conference dinner

Tuesday 21th June: Conference dinner at the restaurant ENTE, Martinsplatz 2a, (close to the Bonn Minster, directly in the city center), starting at 19:00. You will be charged for the dinner directly by the restaurant.

Organisers

- Prof. Markus Hegland
- Prof. Vladimir Pestov
- Prof. Ian Sloan
- Prof. Henryk Woźniakowski

Acknowledgement

The workshop is supported by the Hausdorff Research Institute for Mathematics (HIM) as part of its Trimester Program on Analysis and Numerics for High Dimensional Problems. We thank the HIM-Team for their assistance.
<table>
<thead>
<tr>
<th>Time</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chair</td>
<td>Hegland</td>
<td>Woźniakowski</td>
<td>Gorban</td>
<td></td>
<td>Pestov</td>
</tr>
<tr>
<td>09:30-10:30</td>
<td>Pestov</td>
<td>Kühn</td>
<td>Novak</td>
<td>Merényi</td>
<td>Coffee</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Coffee</td>
</tr>
<tr>
<td>11:15-12:15</td>
<td>Binev</td>
<td>Gnewuch</td>
<td>Wasilkowski</td>
<td>Discussion</td>
<td>Lunch</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lunch</td>
</tr>
<tr>
<td>Chair</td>
<td>Wasilkowski</td>
<td>Novak</td>
<td>Sloan</td>
<td>Hinrichs</td>
<td></td>
</tr>
<tr>
<td>14:15-15:15</td>
<td>Gorban</td>
<td>Hinrichs</td>
<td>Plaskota</td>
<td>Rauhut</td>
<td>Cake</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Cake</td>
</tr>
<tr>
<td>16:00-17:00</td>
<td>Dick</td>
<td>Woźniakowski</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Detailed program

Mon, 09:00-09:30 Hegland, Opening
Mon, 09:30-10:30 Pestov, On the geometry and complexity of similarity search in high dimensions
Mon, 11:15-12:15 Binev, Sparse occupancy trees for approximation and classification
Mon, 14:15-15:15 Gorban, Entropies
Tue, 09:30-10:30 Kühn, Some applications of metric entropy in high-dimensional problems
Tue, 11:15-12:15 Gnewuch, When are multivariate and infinite-dimensional integration on reproducing kernel Hilbert spaces related to geometric discrepancy
Tue, 14:15-15:15 Hinrichs, The curse of dimensionality for monotone functions and for convex functions
Tue, 16:00-17:00 Dick, High dimensional numerical integration
Wed, 09:30-10:30 Novak, Discontinuous information in the worst case and randomized settings
Wed, 11:15-12:15 Wasilkowski, Tractability of $L_2$-approximation for $\infty$-variate functions; liberating the dimension
Wed, 14:15-15:15 Plaskota, Complexity of approximation and integration of piecewise smooth functions
Wed, 16:00-17:00 Woźniakowski, Open tractability problems
Fri, 09:30-10:30 Merényi, Manifold learning for simultaneous inference of multiple latent variables
Fri, 11:15-12:15 Pestov, Discussion
Fri, 14:15-15:15 Rauhut, The Gelfand widths of $l_p$-balls for $0 < p \leq 1$
Authors and Participants

Binev, 3 4 7
Dick, 3 4 12
Gnewuch, 3 4 10
Gorban, 3 4 8
Hegland, 1 4
Hinrichs, 3 4 11
Kühn, 3 4 9
Merényi, 3 4 17
Novak, 3 4 13
Pestov, 1 4 6
Plaskota, 3 4 15
Rauhut, 3 4 19
Sloan, 1 3
Wasilkowski, 3 4 14
Woźniakowski, 3 4 16
On the geometry and complexity of similarity search in high dimensions

Mon, 09:30-10:30

V. Pestov
University of Ottawa, Canada
vpest283@uottawa.ca

Abstract: We will discuss the curse of dimensionality in the context of exact similarity search in metric spaces. The basic questions here are, first, what makes data "high-dimensional", and second, is the curse of dimensionality really inherent in such data? Partial answers are provided by the phenomenon of concentration of measure combined with the Vapnik-Chervonenkis theory of statistical learning. We will also discuss the still open "curse of dimensionality conjecture", as well as an apparent greater efficiency of the approximate nearest neighbour search. The talk is in large part (though not exclusively) based on the preprint arXiv:1008.5105v5 [cs.DS].
Sparse occupancy trees for approximation and classification
Mon, 11:15-12:15

P. Binev
Department of Mathematics
University of South Carolina
Columbia, SC 29208
binev@math.sc.edu

Methods for high-dimensional function approximation and statistical learning are commonly categorized into two classes: parametric and non-parametric.

The parametric methods try to fit the function globally, typically prescribing the structure of the approximant as the combination of a fixed set of trial functions and learning their coefficients by optimizing some error norm. Here one can think, for example, of generalized additive models, projection pursuit, artificial neural networks, or low rank tensor-product approximations. Although these methods have been applied successfully in a large number of applications, they also have some drawbacks. First, the class of functions that are approximated well by such techniques is typically small, and the right model has to be determined a priori. Second, the training stage usually involves the solution of a non-linear optimization problem. This might be a demanding and time-consuming process, thereby effectively limiting the size of the data that can be handled. Furthermore these approximations cannot easily be adjusted to new data, as in the case of incremental online learning or in applications where the domain in which the function is to be evaluated changes over time.

The non-parametric methods try to fit the function locally, usually by partitioning the input space and then using simple local models like piecewise constant approximations. The idea of being content with piecewise constants is supported by classical concentration of measure results according to which a well-behaved function (e.g., Lipschitz-continuous) in very high dimensions deviates much from its mean or median only on sets of small measure. A typical example for such a recovery strategy is to determine for any given query point its $k$ nearest neighbors in the given data and to use their average as the approximate function value. At first glance this kind of memory-based learning does not seem to require any training process except of reading and storing the incoming data. In practice, however, it is necessary to design a data structure that provides a fast solution to the question “what are the nearest neighbors of a query point $x$?” Unfortunately, the exact solution requires either a preprocessing time which is exponential in the dimension $d$ or a single query time which is linear in the number of points $N$. Actually, for function recovery purposes one would also be satisfied with an approximate solution that could be achieved much more efficiently. Therefore, in situations where a fast answer to a query matters alternative strategies may be preferable.

In view of these considerations we develop and investigate some methods, which are primarily designed to be fast and to deal efficiently with large data sets and to provide fast algorithms for evaluation. The motivation behind our approach is to explore the potential of multiresolution ideas for high spatial dimension. For this purpose we propose new methods based on what we call *sparse occupancy trees* and piecewise linear schemes on simplex subdivisions.
Entropy was born in the 19th century as a daughter of energy. Clausius, Boltzmann and Gibbs (and others) developed the physical notion of entropy. In the 20th century, Hartley (1928) and Shannon (1948) introduced a logarithmic measure of information in electronic communication in order “to eliminate the psychological factors involved and to establish a measure of information in terms of purely physical quantities”. Information theory focused on entropy as a measure of uncertainty of subjective choice. This understanding of entropy was returned from information theory to statistical mechanics by Jaynes as a basis of “subjective” statistical mechanics. The entropy maximum approach was declared as a minimization of the subjective uncertainty. This approach gave rise to a MaxEnt “anarchism”. Many new entropies were invented and now there exists rich choice of entropies for fitting needs. The most celebrated of them are the Renyi entropy, the Burg entropy, the Tsallis entropy and the Cressie–Read family. MaxEnt approach is conditional maximization of entropy for the evaluation of the probability distribution when our information is partial and incomplete. The entropy function may be the classical BGS entropy or any function from the rich family of non-classical entropies. This rich choice causes a new problem: which entropy is better for a given class of applications? The MaxEnt “anarchism” was criticized many times as a “senseless fitting”, nevertheless, it remains a very popular approach to multidimensional problems with uncertainty.

We understand entropy as a measure of uncertainty which increases in Markov processes. In this talk, we review the families of non-classical entropies and discuss the question: is entropy a function or an order? We describe the most general ordering of the distribution space, with respect to which all continuous-time Markov processes are monotonic (the *Markov order*). For inference, this approach results in a convex compact set of conditionally “most random” distributions.

Joint work with

P. Gorban (Siberian Federal University, Russia) and
G. Judge (University of California, Berkeley, CA, USA)

References

Some applications of metric entropy in high-dimensional problems

Tue, 09:30-10:30

T. Kühn
Universität Leipzig, Germany
kuehn@math.uni-leipzig.de

Metric entropy plays an important role in many branches of mathematics, e.g., in approximation theory, functional analysis, probability on Banach spaces, and more recently also in the mathematical theory of learning.

In the talk I will discuss some of these applications and present several relevant estimates of entropy and covering numbers in the context of

- finite-dimensional $\ell_p$-spaces
- embeddings of weighted Besov spaces on $\mathbb{R}^d$
- small deviations of Gaussian processes
- Gaussian RKHSs over the $d$-dimensional unit cube.

In particular, it will be discussed how the bounds in these estimates depend on the dimension $d$ (of the sequence spaces, the underlying domain of the function spaces, or the index set of the Gaussian process, respectively).
When are multivariate and infinite-dimensional integration on reproducing kernel Hilbert spaces related to geometric discrepancy

Tue, 11:15-12:15

M. Gnewuch

Institut für Informatik
Christian-Albrechts-Universität zu Kiel
mig@informatik.uni-kiel.de

In many cases multivariate numerical integration is intimately related to geometric discrepancy. Recently, Erich Novak and Henryk Woźniakowski proved a quite general relation between integration on (unweighted) reproducing kernel Hilbert spaces and (unweighted) geometric \( L_2 \)-discrepancy (which they called \( L_2-B \)-discrepancy).

We extend their notion of \( L_2-B \)-discrepancy to \textit{weighted geometric \( L_2 \)-discrepancy}. This extension allows for weights in order to moderate the importance of different groups of variables, as well as for volume measures different from the Lebesgue measure and classes of test sets different from subsets of Euclidean spaces.

We relate the weighted geometric \( L_2 \)-discrepancy to numerical integration defined over weighted reproducing kernel Hilbert spaces and settle in this way an open problem posed by Novak and Woźniakowski in their recent book.

Furthermore, we prove an upper bound for the numerical integration error for cubature formulas that use admissible sample points. The set of admissible sample points may actually be a subset of the integration domain of measure zero. Particularly in infinite-dimensional numerical integration it is important to distinguish between the whole integration domain and the set of those sample points that actually can be used by the algorithms.

The main example to illustrate the (more abstract) results will be infinite-dimensional integration.

References


We study the integration and approximation problems for monotone or convex bounded functions that depend on \( d \) variables, where \( d \) can be arbitrarily large. We consider the worst case error for algorithms that use finitely many function values. We prove that these problems suffer from the curse of dimensionality. That is, one needs exponentially many (in \( d \)) function values to achieve an error \( \varepsilon \).

This is joint work with Erich Novak and Henryk Woźniakowski.
High dimensional numerical integration

Tue, 16:00-17:00

J. Dick
School of Mathematics and Statistics
University of New South Wales
Sydney, NSW, 2052
Australia
josef.dick@unsw.edu.au

In this talk we give an overview of quasi-Monte Carlo rules, which are quadrature rules
\[ \frac{1}{N} \sum_{n=1}^{N} f(y_n) \approx \int_{[0,1]^s} f(x) \, dx. \]  
We focus on the case where the quadrature points are obtained from a so-called digital net. We describe rules for which one obtains integration errors for functions \( f : [0,1]^s \to \mathbb{R} \) of smoothness \( \alpha \geq 1 \) of the form \( O(N^{-\alpha}(\log N)^s) \). We also consider so-called polynomial lattice point sets, which are a special case of digital nets. We show that for this class of quadrature rules one can obtain error bounds for weighted function spaces (with weights \( \gamma = (\gamma_1, \gamma_2, \ldots) \)), which achieve an order of convergence \( C_{\tau,\gamma} N^{-\tau} \) for all \( \tau < \alpha \), where the constant \( C_{\delta,\gamma} > 0 \) is independent of the dimension if the weights satisfy \( \sum_{i=1}^{\infty} \gamma_i^{1/\tau} < \infty \).
Wednesday, 22.06.2011

Discontinuous information in the worst case and randomized settings

Wed, 09:30-10:30

E. Novak
Mathematisches Institut, Universität Jena
Ernst-Abbe-Platz 2, D-07740 Jena, Germany
ERICH.NOVAK@UNI-JENA.DE

We believe that discontinuous linear information is never more powerful than continuous linear information for approximating continuous operators. We prove such a result in the worst case setting. In the randomized setting we consider compact linear operators defined between Hilbert spaces. In this case, the use of discontinuous linear information in the randomized setting cannot be much more powerful than continuous linear information in the worst case setting. These results can be applied when function evaluations are used even if function values are defined only almost everywhere.

Joint work with Aicke Hinrichs and Henryk Woźniakowski.
Tractability of $L_2$-approximation for $\infty$-variate functions; liberating the dimension

Wed, 11:15-12:15

G. W. Wasilkowski
University of Kentucky, USA
greg@cs.engr.uky.edu

The majority of papers on the complexity of multivariate problems consider spaces of functions with $d$ variables for finite yet arbitrarily large $d$. A typical question addressed in these papers is whether the problem is tractable, i.e., whether the minimal number of linear functionals or function values needed to obtain an approximation with error at most $\varepsilon$ is not exponential in $1/\varepsilon$ and $d$. There are many positive results. However, since $d$ can be arbitrarily large independently of $\varepsilon$, there are also many negative results.

There are important problems that deal with functions of infinitely many variables, e.g., path integration. Such problems can be approximated by problems with finite $d$; however, this may result in less efficient methods. In this talk, we discuss recent results on complexity and tractability of weighted $L_2$ approximation of infinitely variate functions. In this sense we liberate the dimension.

We assume that functions $f$ are infinite combinations of functions of finitely many variables and the information about specific $f$ is provided by a finite number of values of linear functionals or function samples, each with finitely many active variables. For example, if we want to compute $f(x)$ then active variables may be defined as nonzero components of $x$.

It turns out that, under relatively natural assumptions, efficient algorithms use information functionals with only $o(\ln(1/\varepsilon))$ of active variables. This implies tractability even when the cost function $\$ is exponential or super exponential.
Complexity of approximation and integration of piecewise smooth functions

Wed, 14:15-15:15

L. Plaskota
Institute of Applied Mathematics and Mechanics
Faculty of Mathematics, Informatics, and Mechanics
University of Warsaw
LESZEKP@MIMUW.EDU.PL

We present recent results on numerical approximation and integration of functions $f$ that are piecewise smooth and singularities are unknown. The approximations are constructed based on finitely many samples of $f$. We consider functions of one variable with point singularities, and functions of two variables with singularities along a smooth curve.

When functions are globally smooth, optimal algorithms are nonadaptive. For piecewise smooth functions any nonadaptive algorithm must fail since it is unable to successfully locate singularities. Can we do better by using adaptive algorithms? We show that the answer is often positive. Moreover, we show that in many cases best adaptive algorithms achieve the same level of accuracy as best algorithms for globally smooth functions.

Apart from the new practical algorithms for piecewise smooth functions, the obtained results provide new insight into the problem of adaption versus nonadaption in the worst case and asymptotic settings within the theory of information-based complexity.

This is a joint work with Greg Wasilkowski, Yaxi Zhao, and Krzysztof Kowalak.
Open tractability problems

Wed, 16:00-17:00

H. Woźniakowski Columbia University and University of Warsaw
HENRYK@CS.COLUMBIA.EDU

In our book with Erich Novak "Tractability of Multivariate Problems" we identified more than 100 open problems related to various tractability questions in different settings. In this talk, we discuss three of them covering the worst case, average case and randomized settings. We believe that these problems are characteristic and properly summarize what is known and what is still open for tractability study.
Friday, 24.06.2011

Manifold learning for simultaneous inference of multiple latent variables
Fri, 09:30-10:30

E. Merényi
Department of Statistics
Department of Electrical and Computer Engineering
Rice University MS 138, 6100 Main Street, Houston, TX 77005 USA
ERZSEBET@RICE.EDU

This talk will present an approach to inferring the values of multiple latent variables which have highly non-linear dependencies on the observable variables, and also exhibit non-linear entanglement with one another.

The problem is motivated – among other applications – by remotely observed, high-dimensional spectroscopic data of planetary surfaces (e.g., reflected sunlight measured at \( n \) different wavelengths for each location of interest, where \( n \) can be hundreds or thousands). Some physical quantities of primary importance for science, such as surface chemical composition, temperature, particle size, cannot be measured directly from spacecraft flying by or orbiting a planetary body, but they can be inferred – in principle – from the observed spectra on which each of them has some (different) global influence. However, traditional – physics based - forward modeling or regression schemes yield only partial results.

We approach this challenge by prototype based manifold learning that directly builds the Delaunay triangulation - thus the Voronoi cells - with respect to the given prototypes of the \( n \)-dimensional data space (based on Martinetz and Schulten, 1994). Once the prototypes are learned to satisfactorily represent the data distribution we evaluate the relationships of the neighboring Voronoi centroids (neighboring \( n \)-dimensional prototypes). This evaluation quantifies and ranks their contributions to sharing the representation of the latent variables \( l_i \) \( (i = 1, \ldots, L) \), and it indicates that accurate inference of different \( l_i \) requires recovery of information from a different number, \( k_i \), of the most important Voronoi neighbor prototypes. We give a principled way to determine \( k_i \) for each \( l_i \) based on a connectivity measure that exploits the anisotropy of the data distribution in the Voronoi cells (Ta¸sdemir and Merényi, 2009, Merényi et al., 2009). Training a simple perceptron with the relative contribution strengths of the top ranking \( k_i \) prototypes corresponding to each \( l_i \) yields highly accurate inferences, in contrast to using information from the same number of prototypes for all \( l_i \) (Zhang and Merényi 2010, Zhang et al., 2010).

References


The Gelfand widths of $\ell_p$-balls for $0 < p \leq 1$

Fri, 14:15-15:15

H. Rauhut
Hausdorff Center for Mathematics
Endenicher Allee 60
53115 Bonn
Germany
RAUHUT@HCM.UNI-BONN.DE

Gelfand widths are closely related to the central problem in information based complexity of providing error estimates of best possible measurement and reconstruction schemes. In the context of compressive sensing (sparse recovery) estimating the performance on $\ell_p$-balls, $0 < p \leq 1$, is of crucial importance because vectors in such balls can be well-approximated by sparse ones. We provide sharp lower and upper bounds for the Gelfand widths of $\ell_p$-balls in the $N$-dimensional $\ell^N_q$-space for $0 < p \leq 1$ and $p < q \leq 2$. Our proofs rely on methods from compressive sensing.

Joint work with S. Foucart, A. Pajor, T. Ullrich
Notes