Report from Dagstuhl Seminar 11051

# Sparse Representations and Efficient Sensing of Data


**Edited by**

**Stephan Dahlke[1], Michael Elad[2], Yonina Eldar[3], Gitta Kutyniok[4], and Gerd Teschke[5]**

1 **Universität Marburg, DE,** dahlke@Mathematik.Uni-Marburg.de
2 **Technion – Haifa, IL,** elad@cs.technion.ac.il
3 **Technion – Haifa, IL,** yonina@ee.technion.ac.il
4 **Universität Osnabrück, DE,** kutyniok@math.uni-osnabrueck.de
5 **Hochschule Neubrandenburg, DE,** teschke@hs-nb.de



## Abstract

This report documents the program and the outcomes of Dagstuhl Seminar 11051 "Sparse Representations and Efficient Sensing of Data". The scope of the seminar was twofold. First, we wanted to elaborate the state of the art in the field of sparse data representation and corresponding efficient data sensing methods. Second, we planned to explore and analyze the impact of methods in computational science disciplines that serve these fields, and the possible resources allocated for industrial applications.



**Seminar** 30. January–04. February, 2011 – www.dagstuhl.de/11051

**1998 ACM Subject Classification** F.2 Analysis of Algorithms and Problem Complexity, G.1 Numerical Analysis, I.4 Image Processing and Computer Vision, J.2 Physical Sciences and Engineering

**Keywords and phrases** Efficient signal sensing schemes, sparse signal representations, efficient signal reconstruction algorithms, impact of the methods in neighboring research fields and applications.

**Digital Object Identifier** 10.4230/DagRep.1.1.108


## 1 Executive Summary

*Stephan Dahlke*
*Michael Elad*
*Yonina Eldar*
*Gitta Kutyniok*
*Gerd Teschke*



Modeling of data is the crucial point in enabling various processing of it. This modeling can take many forms and shapes: it can be done in a low-level way that ties the data samples directly or in higher levels that search for structures and constellations. The task of modeling data is so fundamental that it is underlying most of the major achievements in the fields of signal and image processing. This is true also for processing of more general data sources. Indeed, the field of machine learning that addresses this general problem also recognizes the importance of such modeling. In this realm of models, there is one that stands out as quite simple yet very important this is a model based on sparse description of the data. The core idea is to consider the data as a sparse linear combination of core elements, referred to as atoms. This model has attracted huge interest in the past decade, with many mathematicians, computer scientists, engineers, and scientists from various disciplines working on its different facets, and building a set of tools that lead all the way from pure mathematical concepts to practical tools to be used in other computational sciences as well as applications. Using this model, researchers have shown in recent years a wide battery of computational research disciplines and applications that directly benefit from it, leading to state-of-the- art results. Various reconstruction problems, data compression, sampling and sensing, separation of signals, cleaning and purifying data, adaptive numerical schemes, and more, all require the utilization of sparse representations to succeed in their tasks.

The goals of the seminar can be summarized as follows:
- Establish communication between different focusses of research
- Open new areas of applications
- Manifest the future direction of the field
- Introduce young scientists

To reach these seminar goals, the organizers identified in advance the most relevant fields of research:
- Sampling and Compressed Sensing
- Frames, Adaptivity and Stability
- Algorithms and Applications

The seminar was mainly centered around these topics, and the talks and discussion groups were clustered accordingly. During the seminar, it has turned out that in particular 'generalized sensing', 'data modeling', and corresponding 'algorithms' are currently the most important topics. Indeed, most of the proposed talks were concerned with these three issues. This finding was also manifested by the discussion groups. For a detailed description of the outcome of the discussion, we refer to Section 4.

The course of the seminar gave the impression that sparsity with all its facets is definitely one of the most important techniques in applied mathematics and computer sciences. Also of great importance are associated sampling issues. We have seen many different view points ranging from classical linear and nonlinear to compressive sensing. In particular, new results on generalized sampling show how to design effective sampling strategies for recovering sparse signals. The impact of these techniques became clear as they allow an extension of the classical finite dimensional theory of compressive sensing to infinite dimensional data models. Moreover, it was fascinating to see how sampling and sparsity concepts are by now influencing many different fields of applications ranging from image processing / compression / resolution to adaptive numerical schemes and the treatment of operator equations/inverse problems. It seems that the duality between sparse sampling and sparse recovery is a common fundamental structure behind many different applications. However, the mathematical technicalities remain quite challenging. As algorithmic issues were also discussed quite intensively, we could figure out that we are now essentially at some point where $\ell_1$-optimization is competitive speed-wise with classical linear methods such as conjugate gradient.

Summarizing our findings during the seminar, we believe that the research agenda can be more focused on the actual bottlenecks, being in problem/signal modeling, design of sampling and recovery methods adapted to specific problems, and algorithmic improvements including performance bounds and guarantees.

## 2 Table of Contents

# 3 Overview of Talks

## 3.1 Multivariate Periodic Function Spaces

*Ronny Bergmann (Universität zu Lübeck, DE)*


**Joint work of** Bergmann, Ronny; Prestin, Jürgen

In this talk we present patterns $\mathcal{P}(\mathbf{M})$ based on invertible matrices $\mathbf{M} \in \mathbb{Z}^{d\times d}$ as multivariate generalizations of the equidistant points on $[0, 1)$. We present different properties of the patterns, e.g. the classification of subpatterns and a dual group, that is used to define a discrete Fourier transform with respect to $\mathbf{M}$.

Using the pattern to generate translates of a square integrable function $f$ defined on the $d$-dimensional $2\pi$-periodic torus, we introduce a translation invariant space $V^f$. This space can be characterized by the Fourier series of $f$ and the aforementioned Fourier transform on $\mathcal{P}(\mathbf{M})$. The same is possible for subspaces that are generated by different translates of $f$ or even translates of other functions in $V^f$. Finally this yields properties for a decomposition of $V^f$ into two orthogonal subspaces that are translation invariant with respect to a certain subpattern of $\mathcal{P}(\mathbf{M})$.

## 3.2 Random tight frames and applications

*Martin Ehler (National Institutes of Health – Bethesda, US)*


**Joint work of** Ehler, Martin; Galanis, Jennifer
**Main reference** M. Ehler, J. Galanis, "Frame theory in directional statistics," Stat. Probabil. Lett., in press.
**URL** http://dx.doi.org/10.1016/j.spl.2011.02.027

It is known that independent, uniformly distributed points on the sphere approximately form a finite unit norm tight frame. In this talk, we introduce probabilistic frames to more deeply study finite frames whose elements are chosen at random. In fact, points chosen from any probabilistic tight frame approximately form a finite tight frame; they do not have to be uniformly distributed, nor have unit norm. We also observe that classes of random matrices used in compressed sensing are induced by probabilistic tight frames.

Finally, we merge directional statistics with frame theory to elucidate directional statistical testing. Distinguishing between uniform and non-uniform sample distributions is a common problem in directional data analysis; however for many tests, non-uniform distributions exist that fail uniformity rejection. We find that probabilistic tight frames yield non-uniform distributions that minimize directional potentials, leading to failure of uniformity rejection for the Bingham test. We apply our results to model patterns found in granular rod experiments.

Related references are:

### References

1 M. Ehler, "Random tight frames," arXiv:1102.4080v1.
2 M. Ehler, K.A. Okoudjou, "Minimization of the probabilistic p-frame potential," arXiv:1101.0140.

## 3.3 Alternating Direction Optimization for Imaging Inverse Problems with Sparsity-Inducing Regularization

*Mario Figueiredo (TU Lisboa, PT)*


**Joint work of** Figueiredo, Mário; Bioucas-Dias, José; Afonso, Manya

Most modern approaches (regularization-based or Bayesian) to imaging inverse problems lead to optimization problems. These (usually convex) problems have features that place them beyond the reach of off-the-shelf optimization methods and have stimulated a significant amount of research. In particular, the presence of regularizers encouraging sparse solutions imply the non-smoothness of the objective function, which together with its typical very high dimensionality constitutes a challenge.

Examples of this include frame-based regularization (either in an analysis or synthesis formulation), where the classical regularizer involves the $\ell_1$ norm.

This talk will cover our recent work on the application of a class of techniques known as "alternating direction methods" to several imaging inverse problems with frame-based sparsity-inducing regularization, namely: (a) standard image restoration/reconstruction from linear observations with Gaussian noise; (b) image restoration from Poissonian observations; (c) multiplicative noise removal.

In all these cases, the proposed methods come with theoretic convergence guarantees and achieve state-of-the-art speed, as shown in the reported experiments. To further illustrate the flexibility of this class of methods, we show how it can be used to seamlessly address hybrid analysis/synthesis formulations as well as group-norm regularizers (with or without group overlap).

The work described in this talk was co-authored by José M. Bioucas-Dias and Manya V. Afonso, and reported in the following publications:

### References

1 M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems", *IEEE Transactions on Image Processing*, vol. 20, pp. 681–695, 2011.
2 M. Afonso, J. Bioucas-Dias, and M. Figueiredo, "Fast image recovery using variable splitting and constrained optimization", *IEEE Transactions on Image Processing*, vol. 19, pp. 2345–2356, 2010.
3 J. Bioucas-Dias, and M. Figueiredo, "Multiplicative noise removal using variable splitting and constrained optimization", *IEEE Transactions on Image Processing*, vol. 19, pp. 1720–1730, 2010.
4 M. Figueiredo and J. Bioucas-Dias "Restoration of Poissonian images using alternating direction optimization", *IEEE Transactions on Image Processing*, vol. 19, pp. 3133–3145, 2010.

## 3.4 Adaptive wavelet methods for inverse parabolic problems

*Ulrich Friedrich (University of Marburg, DE)*



We combine adaptive wavelet techniques for well-posed problems and regularization theory for inverse problems. We are concerned with identifying certain parameters in a parabolic reaction-diffusion equation from measured data. The PDE describes the gene concentrations in embryos at an early state of development. The forward problem is formulated as an evolution equation, and the analytical properties of the parameter-to-state operator are analyzed. The results justify the application of an iterated soft-shrinkage algorithm within a Tikhonov regularization approach. The forward problem is treated by means of a new adaptive wavelet algorithm which is based on tensor wavelets. A generalized anisotropic tensor wavelet basis dealing with complicated domains is given. This leads to dimension independent convergence rates. An implementation of this procedure involving the new adaptive wavelet solver is proposed and numerical results are presented.

## 3.5 A Majorization-minimization algorithm for sparse factorization and some related applications

*Onur G. Guleryuz (DoCoMo USA Labs – Palo Alto, US)*


**Joint work of** Guleryuz, Onur G.; Chou, Evan; Zujovic, Jana

Given a data matrix, $X$, we are interested in its approximations of the form $\hat{X} = TC$ where $T$ and $C$ are two sparse matrices. The optimization setup is

$$\min_{T,C} ||X - TC||_2 \text{ subject to } ||T||_0 + ||C||_0 \leq \kappa.$$

The problem arises in

**(a)** accelerating matrix multiplication where an a priori known X is to be multiplied with a dense matrix S, with S only available online.

**(b)** approximation where $X = Y + W$, with $Y = TC$, and the goal is to recover $Y$ (i.e., structured signal under noise, structured signal with missing data, etc.)

**(c)** compression, with $T$ as the data-adaptive basis, $C$ as the matrix of coefficient vectors, and assuming one is using a nonlinear-approximation-based compression algorithm.

$X$ is specified in a domain spanned by two known orthonormal matrices which we call the presentation basis. Many disciplines approach these applications using the SVD of X, disregarding the last twenty-plus years of research. Our aim is to move toward "generic signal processing" where DSP techniques can be used to match/better existing results with minor or no domain-specific information.

### References

1 M. Döhler, S. Kunis, and D. Potts. Nonequispaced hyperbolic cross fast Fourier transform. *SIAM J. Numer. Anal.*, 47:4415 – 4428, 2010.

2 M. Aharon, M. Elad, and A.M. Bruckstein, The K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation, *IEEE Trans. On Signal Processing* **54(11)** (2006), 4311-4322.

3 J. Zujovic and O. G. Guleryuz, Complexity Regularized Pattern Matching, *Proc. IEEE Int'l Conf. on Image Proc. (ICIP2009), Cairo, Egypt*, Nov. 2009.
4 Rubinstein, R., Zibulevsky, M. and Elad, M., Double Sparsity: Learning Sparse Dictionaries for Sparse Signal Approximation, *IEEE Transactions on Signal Processing* **58(3)** (2010).
5 E. Chou and O. G. Guleryuz, A Majorization-minimization algorithm for sparse factorization and related applications, (in preparation).
6 O. G. Guleryuz, Nonlinear Approximation Based Image Recovery Using Adaptive Sparse Reconstructions and Iterated Denoising: Parts I and II, *IEEE Transactions on Image Processing*, (2006).


## 3.6 Stable discretizations of the hyperbolic cross fast Fourier transform

*Stefan Kunis (Universität Osnabrück, DE)*


**Joint work of** Kämmerer, Lutz; Kunis, Stefan; Potts, Daniel

A straightforward discretization of problems in $d$ spatial dimensions with $2^n$, $n \in \mathbb{N}$, grid points in each coordinate leads to an exponential growth $2^{dn}$ in the number of degrees of freedom.

We restrict the frequency domain to the hyperbolic cross

$$H_n^d = \bigcup_{\mathbf{j}\in\mathbb{N}_0^d, \|\mathbf{j}\|_1 = n} (-2^{j_1-1}, 2^{j_1-1}] \times \ldots \times (-2^{j_d-1}, 2^{j_d-1}] \cap \mathbb{Z}^d,$$

and ask for the fast approximate evaluation of the trigonometric polynomial

$$f(\mathbf{x}) = \sum_{\mathbf{k}\in H_n^d} \hat{f}_{\mathbf{k}} \, \mathrm{e}^{2\pi \mathrm{i} \mathbf{k}\mathbf{x}}, \tag{1}$$

at nodes $\mathbf{x}_\ell \in \mathbb{T}^d$, $\ell = 1, \ldots, M$.

We note that the reduced problem size is $c_d 2^n n^{d-1} \leq |H_n^d| \leq C_d 2^n n^{d-1}$ and a classical result states the computation of (1) for all sparse grid nodes takes at most $C_d 2^n n^d$ floating point operations.

This has been generalized for arbitrary spatial sampling nodes and both algorithms are available in the Matlab toolbox `nhcfft`.

▶ **Theorem 1.** *[1] The computation of* (1) *at all nodes* $\mathbf{x}_\ell \in \mathbb{T}^d$*,* $\ell = 1, \ldots, |H_n^d|$*, takes at most* $C_d 2^n n^{2d-2}(|\log \varepsilon| + \log n)^d$*, where* $\varepsilon > 0$ *denotes the target accuracy.*

More recently, we analyzed the numerical stability of these sampling sets and in sharp contrast to the ordinary FFT which is unitary, we found the following negative result.

▶ **Theorem 2.** *[2] The computation of* (1) *at the sparse grid has condition number*

$$c_d 2^{\frac{n}{2}} n^{\frac{2d-3}{2}} \leq \kappa \leq C_d 2^{\frac{n}{2}} n^{2d-2}.$$

Although random sampling offers a stable spatial discretization with high probability if $M \geq C|H_n^d| \log |H_n^d|$, the fast algorithm [1] relies on an oversampled sparse grid and thus suffers from the same instability.

Ongoing work [3] considers lattices as spatial discretization for the hyperbolic cross fast Fourier transform. These turn out to have quite large cardinality asymptotically but offer perfect stability and outperform known algorithms by at least one order of magnitude with respect to CPU timings for moderate problem sizes.

**References**

1 M. Döhler, S. Kunis, and D. Potts. Nonequispaced hyperbolic cross fast Fourier transform. *SIAM J. Numer. Anal.*, 47:4415 – 4428, 2010.

2 L. Kämmerer and S. Kunis. On the stability of the hyperbolic cross discrete Fourier transform. *Numer. Math.*, 117:581 – 600, 2011.

3 L. Kämmerer, S. Kunis, and D. Potts. Interpolation lattices for hyperbolic cross trigonometric polynomials. *Preprint 2010-13, Fakultät für Mathematik, TU Chemnitz*, 2010.

## 3.7 Optimally Sparse Image Approximations Using Compactly Supported Shearlets

*Wang-Q Lim (Universität Osnabrück, DE)*



It is now widely acknowledged that analyzing the intrinsic geometrical features of a function/signal is essential in many applications. In order to achieve this, several directional systems have been proposed in the past. The first breakthrough was achieved by Candes and Donoho who introduced curvelets and showed that curvelets provide an optimal approximation property for a special class of 2D piecewise smooth functions, called cartoon-like images. However, only band-limited directional systems providing an optimal approximation property have been constructed so far, except adaptive representation schemes.

In this talk, we will show that optimally sparse approximation of cartoon-like images can be achieved using compactly supported shearlet frames in both 2D and 3D. We then briefly discuss our ongoing work to construct a compactly supported directional system which is not only a tight frame but also provides optimally sparse approximation of cartoon-like images.

## 3.8 Exact test instances for Basis Pursuit Denoising

*Dirk Lorenz (TU Braunschweig, DE)*


**Main reference** Dirk A. Lorenz, "Constructing test instances for Basis Pursuit Denoising," submitted for publication, 2011.
**URL** http://arxiv.org/abs/1103.2897

The number of available algorithms for the so-called Basis Pursuit Denoising problem (or the related LASSO-problem) is large and keeps growing. Similarly, the number of experiments to evaluate and compare these algorithms on different instances is growing. However, many comparisons lack of test instances for which exact solutions are known. We propose to close this gap by a procedure which calculates a right hand side from a given matrix, regularization parameter and a given solution. It can be shown that this can be accomplished by means of projection onto convex sets (POCS) or quadratic programming. The method has been implemented in MATLAB and is available as part of L1TestPack from http://www.tu-braunschweig.de/iaa/personal/lorenz/l1testpack.

## 3.9 Cosparse Analysis Modeling – Uniqueness and Algorithms

*Sangnam Nam (INRIA – Rennes, FR)*



In the past decade there has been a great interest in a synthesis-based model for signals, based on sparse and redundant representations. Such a model assumes that the signal of interest can be composed as a linear combination of few columns from a given matrix (the dictionary). An alternative analysis-based model can be envisioned, where an analysis operator multiplies the signal, leading to a cosparse outcome. In this work, we consider this analysis model, in the context of a generic missing data problem (e.g., compressed sensing, inpainting, source separation, etc.). Our work proposes a uniqueness result for the solution of this problem, based on properties of the analysis operator and the measurement matrix. We also considers two algorithms for solving the missing data problem, an L1-based and a new greedy method. Our simulations demonstrate the appeal of the analysis model, and the success of the pursuit techniques presented.

**References**

1 M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inverse Problems*, vol. 23, no. 3, pp. 947–968, June 2007.
2 I. W. Selesnick and M. A. T. Figueiredo, "Signal restoration with overcomplete wavelet transforms: Comparison of analysis and synthesis priors," *In Proceedings of SPIE*, vol. 7446 (Wavelets XIII), August 2009.
3 E. J. Candès, Y. C. Eldar, D. Needell, and Y. Ma, "Compressed sensing with coherent and redundant dictionaries," Tech. Rep. 1005.2613v1, arXiv, 2010.
4 J. Portilla, "Image restoration through l0 analysis-based sparse optimization in tight frames," *Proceedings of the 16th IEEE International Conference on Image Processing*, pp. 3865–3868, 2009.
5 R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.
6 D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell^1$ minimization," *Proc. Nat. Aca. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.
7 Y. M. Lu and M. N. Do, "Sampling signals from a union of subspaces," *IEEE Signal Processing Magazine*, pp. 41–47, mar 2008.
8 T. Blumensath and M. E. Davies, "Sampling theorems for signals from the union of finite-dimensional linear subspaces," *IEEE Transactions on Information Theory*, vol. 55, no. 4, pp. 1872–1882, 2009.
9 M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab Software for Disciplined Convex Programming," http://cvxr.com/cvx, August 2008.

## 3.10 Space Splittings and Schwarz-Southwell Iterations

*Peter Oswald (Jacobs University – Bremen, DE)*


**Joint work of** Griebel, Michael; Oswald, Peter

The talk reviews joint work in progress with M. Griebel [1], which continues our collaboration on iterative solvers for symmetric elliptic variational problems that are based on Hilbert space splittings, so-called additive and multiplicative Schwarz methods. Note that Hilbert space splittings underlying the theory of Schwarz methods have lately reappeared as fusion frames. See [2] for a review of previous work.

While in the standard theory of multiplicative Schwarz methods the order of subproblem traversal is fixed, in the new versions the ordering is chosen in a weak greedy fashion, e.g., according to the size of subproblem residuals, or randomly. For linear systems and Gauss-Seidel methods (a special instance of multiplicative Schwarz methods) the greedy ordering goes back to Gauss and Seidel, and has been popularized by Southwell in the 1940-50ies. The method has been theoretically analyzed in the framework of coordinate descent methods for convex optimization methods, and has lately been revived in the context of sparse approximation.

Given these developments, we decided to first formulate and prove convergence results for Schwarz-Southwell methods for the case of splittings into $N$ subproblems. The main result is an exponential energy error decay estimate of the form

$$\|u - u^{(m+1)}\|_E^2 \le (1 - \frac{\gamma}{N})\|u - u^{(m)}\|_E^2, \qquad m \ge 0,$$

where $\gamma$ depends on the spectral bounds characterizing the space splitting, the relaxation parameter $\omega$, and the weakness parameter $\beta$ of the weak greedy step. The result shows that greedy strategies can slightly improve the performance of multiplicative Schwarz methods. We also state a similar estimate for the expected convergence rate if the subproblem ordering is randomized. Investigations on infinite splittings are still at their beginning, they benefit from the theory of greedy algorithms in infinite-dimensional Hilbert and Banach spaces developed by Temlyakov and others. We hope that a better understanding of this topic will shed new light on adaptive multilevel methods such as the early work by Rüde.

### References

1 M. Griebel, P. Oswald, *Schwarz-Southwell Methods*. In preparation.
2 P. Oswald, *Stable space splittings and fusion frames*. In Wavelets XIII (V.K. Goyal, M. Papadakis, D. Van de Ville, eds.), Proc. SPIE Vol. 7446 (SPIE, Bellingham, 2009), 744611.

## 3.11 Sparse Approximation of Images by the Easy Path Wavelet Transform

*Gerlind Plonka-Hoch (Universität Göttingen, DE)*



The Easy Path Wavelet Transform (EPWT) has recently been proposed as a tool for sparse representations of bivariate functions from discrete data, in particular from image data.

The EPWT is a locally adaptive wavelet transform. It works along pathways through the array of function values and it exploits the local correlations of the given data in a simple appropriate manner.

Using polyharmonic spline interpolation, we show that the EPWT leads, for a suitable choice of the pathways, to optimal $N$-term approximations for piecewise Hölder smooth functions with singularities along curves.

## 3.12 Quadrature errors, discrepancies and variational dithering

*Gabriele Steidl (Universität Kaiserslautern, DE)*


**Main reference** M. Gräf, D. Potts and G. Steidl, Quadrature errors, discrepancies and their relations to halftoning on the torus and the sphere, Preprint TU Chemnitz, Fakultät für Mathematik, Preprint 5, 2011.
**URL** http://www.mathematik.uni-kl.de/ steidl

The stippling technique places black dots such that their density gives the impression of tone.

The original idea for considering minimizers of this functional as 'good' dot positions comes from electrostatic principles. This talk is related to the continuous version of the above attraction-repulsion functional with more general functions $\varphi : [0,\infty) \to \mathbb{R}$:

$$E_\varphi(p) := \frac{\lambda}{2} \sum_{i,j=1}^{M} \varphi(\|p_i - p_j\|_2) - \sum_{i=1}^{M} \int_{[0,1]^2} w(x)\varphi(\|p_i - x\|_2)\, dx, \tag{2}$$

where $w : [0,1]^2 \to [0,1]$ and $\lambda := \frac{1}{M} \int_{[0,1]^2} w(x)\, dx$. The function $\varphi(r) = -r$ was used in as well as $\varphi(r) = -\log(r)$. In another paper, the authors mentioned $\varphi(r) = -r^\tau$, $0 < \tau < 2$ and $\varphi(r) = r^{-\tau}$, $\tau > 0$ for $r \neq 0$. In this talk we relate stippling processes with the classical mathematical question of finding best nodes for quadrature rules. We provide theoretical results on the connection between seemingly different concepts, namely quadrature rules, attraction-repulsion functionals, $L_2$–discrepancies and least squares functionals. For the later approach we provide numerical minimization algorithms. In the theoretical part, we start with worst case quadrature errors on RKHSs in dependence on the quadrature nodes. While in the literature, this was mainly done for constant weights $w \equiv 1$, see [5], we incorporate a weight function related to the image into the quadrature functional. The corresponding quadrature error $\mathrm{err}_K(p)$ which depends on the reproducing kernel $K$ can be defined for RKHSs on $\mathcal{X} \in \{\mathbb{R}^2, [0,1]^2\}$ as well as for RKHSs on compact manifolds like $\mathcal{X} \in \{T^2, S^2\}$. We aim to minimize this quadrature error in order to obtain optimal quadrature nodes $p$. It turns out that for special kernels $K$ (on special spaces $\mathcal{X}$) this quadrature error (or at least its minimizers) covers the following approaches:

1. Attraction-Repulsion Functionals
   An interesting case of RKHSs appears for radial kernels $K(x,y) = \varphi(\|x - y\|_2)$ depending only on the distance of the points. We will show that in this case the quadrature error $\mathrm{err}_K(p)$ can be considered as a generalization of (2) which works not only on $[0,1]^2$ but also to compact manifolds. Hence our approach goes far beyond the setting in [1] or [2].
2. $L_2$–Discrepancies
   We prove that for $\mathcal{X} \in \{[0,1]^2, T^2, S^2\}$ and discrepancy kernels $K$, the quadrature errors on RKHSs defined by these kernels coincide with $L_2$–discrepancy functionals. For various applications of $L_2$–discrepancy functionals, see [5] and the references therein. Note

that a relation between the distance kernels $K(x, y) = \|x - y\|_2$ on $T^2$ and $S^2$ and the corresponding discrepancy kernels was shown numerically in [3].

3. Least Squares Functionals
   Finally, we consider RKHSs of bandlimited functions with bandlimited kernels on $\mathcal{X} \in \{T^2, S^2\}$. The reason for addressing these spaces is that we want to approximate functions on $\mathcal{X}$ by bandlimited functions in order to apply fast Fourier techniques. We prove that for these RKHSs the quadrature error can be rewritten as a least squares functional.

In the numerical part we approximate functions and kernels on $\mathcal{X} \in \{T^2, S^2\}$ by their bandlimited versions and minimize the corresponding quadrature error which takes in this case the form of a least squares functional. Due to the page limitation we restrict our attention to the sphere $S^2$. We are not aware of any results on $S^2$–stippling in the literature. We propose a nonlinear CG method on manifolds to compute a minimizer of the least squares functional on $S^2$. This method was also successfully used for the approximation of spherical designs, i.e., for $w \sim 1$ in [4] and is generalized in this paper. In particular, each CG step can be realized in an efficient way by the *nonequispaced fast spherical Fourier transform* (NFSFT). This reduces the asymptotic complexity of the proposed algorithm drastically, e.g., from $\mathcal{O}(MN^2)$ to $\mathcal{O}(N^2 \log^2 N + M \log^2(1/\epsilon))$ arithmetic operations per iteration step, where $\epsilon$ is the described accuracy and $N$ corresponds to the bandwidth. In other words, only by the help of the NFSFT the computation becomes possible in a reasonable time.

### References


1 T. Teuber, G. Steidl, P. Gwosdek, Ch. Schmaltz and J. Weickert, Dithering by differences of convex functions, *SIAM J. on Imaging Science* **4(1)**, 79 - 108, 2011.

2 C. Schmaltz, P. Gwosdek, A. Bruhn and J. Weickert, Electrostatic halftoning, *Computer Graphics Forum*, **29**, 2313 - 2327, 2010.

3 M. Gräf, D. Potts and G. Steidl, Quadrature errors, discrepancies and their relations to halftoning on the torus and the sphere, *TU Chemnitz, Fakultät für Mathematik, Preprint 5*, 2011.

4 M. Gräf and D. Potts, On the computation of spherical designs by a new optimization approach based on fast spherical Fourier transforms, *Preprint, Technical University Chemnitz*, 2010.

5 E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems Volume II: Standard Information for Functionals*, Eur. Math. Society, EMS Tracts in Mathematics Vol 12, 2010.


## 3.13 Compressive sensing and inverse problems

*Gerd Teschke (Hochschule Neubrandenburg, DE)*


**Joint work of** Teschke, Gerd; Herrholz, Evelyn
**Main reference** Evelyn Herrholz, Gerd Teschke, "Compressive sensing principles and iterative sparse recovery for inverse and ill-posed problems," Inverse Problems, vol. 26, no. 12, 125012.
**URL** http://dx.doi.org/10.1088/0266-5611/26/12/125012

We shall be concerned with compressive sampling strategies and sparse recovery principles for linear inverse and ill-posed problems. As the main result, we provide compressed measurement models for ill-posed problems and recovery accuracy estimates for sparse approximations of the solution of the underlying inverse problem. The main ingredients are variational formulations that allow the treatment of ill-posed operator equations in the context of compressively sampled data. In particular, we rely on Tikhonov variational and constrained optimization formulations. One essential difference to the classical compressed sensing framework is the incorporation of joint sparsity measures allowing the treatment of infinite dimensional reconstruction spaces. The theoretical results are furnished with a number of numerical experiments.

We rely on a signal model, $X_k = \{x \in X,\, x = \sum_{\ell \in \mathcal{I}, |\mathcal{I}|=k} \sum_{\lambda \in \Lambda} d_{\ell,\lambda} a_{\ell,\lambda},\;\; d \in (\ell_2(\Lambda))^m\}$, in which we assume that the coefficients $d_{\ell,\lambda}$ share a joint sparsity pattern (only $k$ out of $m$ sequences $\{d_{\ell,\lambda}\}_{\lambda\in\Lambda}$ do not vanish). The space $X_k$ can be seen as a union of (shift invariant) subspaces. One approach to recover $d$ was suggested in [Y. C. Eldar. Compressed Sensing of Analog Signals in Shift-Invariant Spaces. IEEE Trans. on Signal Processing, 57(8), 2009.]. We propose an alternative by solving adequate variational problems. The essential idea to tackle the support set recovery problem is to involve the joint sparsity measure $\Psi_{q,r}(d) = (\sum_{\ell=1}^m (\sum_{\lambda\in\Lambda} |d_{\ell,\lambda}|^r)^{\frac{q}{r}})^{\frac{1}{q}}$. This measure promotes a selection of only those indices $\ell \in \{1,\ldots,m\}$ for which $\|\{d_{\ell,\lambda}\}_{\lambda\in\Lambda}\|_{\ell_r(\Lambda)}$ is large enough, i.e. where the size of the coefficients $d_{\ell,\lambda}$ indicates a significant contribution to the representation of $x$. In order to define an adequate variational formulation, we have to introduce a suitable sensing model. Assume the data $y$ are obtained by sensing $Kx$ through $F_s$ (a compressed version of $F_v$), i.e. $y = F_s K x = F_s K F_a^* d = A F_{K^*v} F_a^* d$, where $K$ is an ill-posed but bounded linear operator and $A$ a sensing matrix satisfying a $2k$-RIP with isometry constant $\delta_{2k}$. If the ansatz systems $\Phi_v$ and $\Phi_a$ diagonalize $K$, we can write a noisy measurement scenario as follows

$$y^\delta = (TD)d + z \quad \text{with} \quad \|z\|_{(\ell_2(\Lambda))^m} \le \delta\ ,$$

where $T$ describes the application of $A$ with respect to each $\lambda$ and where $D$ describes the diagonal matrix performing the application of $K$. To derive an approximation $d^*$ to the solution $d$ of the inverse problem, we propose to solve the following constrained optimization problem

$$\min_{d\in B(\Psi_{1,2},R)} \|y^\delta - (TD)d\|^2_{(\ell_2(\Lambda))^p} + \alpha\|d\|^2_{(\ell_2(\Lambda))^m}\ . \tag{3}$$

The minimizing element $d^*$ of (3) is iteratively approximated by

$$d^{n+1} = \mathbb{P}_R\left(D^*T^*(y^\delta - TDd^n)\frac{\gamma^n}{C} + \left(1 - \frac{\alpha\gamma^n}{C}\right)d^n\right)\ . \tag{4}$$

We can provide the following accuracy estimate for $d^*$.

▶ **Theorem 1.** *Assume $R$ was chosen such that the solution $d$ of problem $y = (TD)d$ does not belong to $B(\Psi_{1,2}, R)$ and that $0 \le \delta_{2k} < \frac{(1+\sqrt{2})\kappa^2_{min} - \kappa^2_{max} + \sqrt{2}\alpha}{(1+\sqrt{2})\kappa^2_{min} + \kappa^2_{max}}$. Then the minimizer $d^*$ of (3) satisfies*

$$\|d^* - d\|_{(\ell_2(\Lambda))^m} \le C_0 k^{-1/2}\Psi_{1,2}(d^k - d) + C_1\|L(d^\dagger - d)\|_{(\ell_2(\Lambda))^m} + C_2\delta + C_3\sqrt{\alpha}R\ ,$$

*where the constants $C_0$, $C_1$, $C_2$, and $C_3$ are given explicitly and where $d^k$ denotes the best $k$-row approximation.*

## 3.14 Sampling in the Age of Sparsity

*Martin Vetterli (EPFL – Lausanne, CH)*



Sampling is a central topic not just in signal processing and communications, but in all fields where the world is analog, but computation is digital. This includes sensing, simulating, and rendering the real world.

The question of sampling is very simple: when is there a one-to-one relationship between a continuous-time function and adequately acquired samples of this function? Sampling has a rich history, dating back to Whittaker, Nyquist, Kotelnikov, Shannon and others, and is an active area of contemporary research with fascinating new results.

The classic result of sampling is the one on bandlimited functions, where taking measurements at the Nyquist rate (or twice the maximum bandwidth) is sufficient for perfect reconstruction. These results were extended to shift-invariant subspaces and multiscale spaces during the development of wavelets, as well as in the context of splines.

All these methods are based on subspace structures, and on linear approximation. Recently, non-linear methods have appeared. Non-linear approximation in wavelet spaces has been shown to be a powerful approximation and compression method. This points to the idea that functions that are sparse in a basis (but not necessarily on a fixed subspace) can be represented efficiently.

The idea is even more general than sparsity in a basis, as pointed out in the framework of signals with finite rate of innovation. Such signals are non-bandlimited continuous-time signals, but with a parametric representation having a finite number of degrees of freedom per unit of time. This leads to sharp results on sampling and reconstruction of such sparse continuous-time signals, namely that 2K measurements are necessary and sufficient to perfectly reconstruct a K-sparse continuous-time signal. In accordance with the principle of parsimony, we call this sampling at Occam's rate. We indicate an order $K^3$ algorithm for reconstruction, and describe the solution when noise is present, or the model is only approximately true.

Next, we consider the connection to compressed sensing and compressive sampling, a recent approach involving random measurement matrices. This is a discrete time, finite dimensional set up, with strong results on possible recovery by relaxing the $\ell_0$ into $\ell_1$ optimization, or using greedy algorithms.

These methods have the advantage of unstructured measurement matrices (actually, typically random ones) and therefore a certain universality, at the cost of some redundancy. We compare the two approaches, highlighting differences, similarities, and respective advantages.

Finally, we move to applications of these results, which cover wideband communications, noise removal, distributed sampling, and super-resolution imaging, to name a few. In particular, we describe a recent result on multichannel sampling with unknown shifts, which leads to an efficient super-resolution imaging method.

## 3.15 Digital Shearlet Transform on Pseudo-Polar Grids

*Xiaosheng Zhuang (Universität Osnabrück, DE)*



We developed a rationally designed digital shearlet theory which is the digitization of the existing shearlet theory for continuum data. Our implementation of the digital shearlet transform is based on utilization of the pseudo-polar Fourier transform, which provide a natural implementation for digital shearlets on the discrete image domain. The pseudo-polar Fourier transform without weighting is generally not an isometry. Isometry can be achieved by careful weighting of the pseudo-polar grid, yet it is difficult to obtain such a weight function.

We showed how efficient weight functions can be designed and obtained on the pseudo-polar grids so that almost isometry can be achieved. In addition, we discussed the software package ShearLab that implements the digital shearlet transform. The ShearLab provides various quantitative measures allowing one to tune parameters and objectively improve the implementation as well as compare different directional transform implementations.

## 4 Discussion Groups and Further Challenges

During the seminar we had three discussion groups. The outcome shall be briefly reviewed in the following three subsections.

### 4.1 Sampling and Compressed Sensing

*Martin Vetterli (EPFL – Lausanne, CH)*
*Hans-Georg Feichtinger (University of Vienna, AT)*



A consensus emerged that there is a fundamental difference between discrete, compressed sensing problems, and continuous, sampling problems. Ignoring this difference will lead to low performance. Several talks highlighted this in the workshop (e.g. Hansen, etc.). The issue of 'designer matrices' for compressed sensing split the audience. On the one hand, some people (e.g. Pfander) said they did not want another discussion on what are good sensing matrices, while others, more concerned about applications, pointed out the importance of fixed (deterministic) and structured (fast) sensing matrices. Random but cyclic matrices where mentioned as a possible alternative. The importance of looking at infinite dimensional problems was reiterated, as well as modeling the underlying physics correctly. This comes into play both in acquisition or analysis (inc. astronomy and parameter estimation) and simulation or synthesis. In sampling and compressed sensing, the role of adaptivity was pointed out as an open problem by Teschke and there was an agreement, even if efficient and/or practical methods have yet to be found. The link to information based complexity was made, since this is a general framework for function class approximation from samples. However, this is a non-constructive theory. Learning problems, be it dictionary learning for compressed sensing or smoothing kernels (point spread functions) for sampling, play also an important role for sparse methods to become practical.

In sum, a certain maturity has been reached, and it is much more clear where compressed sensing or sparse sampling can be used and when not. The research agenda can thus be more focused on the actual bottlenecks, being in modeling, design of methods adapted to specific problems, or algorithmic improvements including performance bounds and guarantees.

## 4.2 Frames, Adaptivity and Stability

*Rob Stevenson (University of Amsterdam, NL)*
*Peter Oswald (Jacobs University - Bremen, DE)*



M. Fornasier pointed to adaptive methods for frame expansions and mentioned to consider hybrid frames as redundancy by itself does not improve the rate of convergence. There is an requirement of using additional structure for obtaining faster convergence. A basic idea could be to use principles from compressed sensing, i.e. use union of bases from sufficiently diverse types, e.g smooth and non-smooth, quarkonians. The questions arises how to define optimality. P. Oswald answered that there can't be a proper definition of optimality besides checking for particular examples. The idea came up to add another parameter for redundancy, i.e. just enrich existing frames and use a measure for mutual in-/coherence. M. Fornasier reminds of work of Bruno Torresani (acoustic signal of 'Glockenspiel' including a transient): when a hammer attacks, followed by a harmonic, best N-term approximation needs bases of different types, and alternating optimization with respect to these different bases, i.e. algorithmic issues have to be taken into account. R. Stevenson reported on numerical tests for Laplace problems using an adaptive scheme (effect of Richardson extrapolation, results depend on sparsity structure on coarse scales). P. Oswald asked the question differently: don't ask which frames (union of two bases) should be used, but rather ask: given a frame, which signals can be approximated efficiently. M. Fornasier mentioned that the typical criterion for choosing the next building block is based on checking the residual, which is justified in case of wavelets and other bases due to incoherency. P. Oswald answered that just looking at incoherence doesn't solve the problem, it just reports numerical examples. U. Friedrich said that the choice of 'different' bases is application driven; he also asked whether there is a theoretical founded approach, which also allows to prove optimal approximation rates? P. Maaß answered: for convergence rates for solving inverse problems one requires a source condition of type $A^w \in \partial R_p(u^\dagger)$ has to be satisfied, i.e. the choice of the frame has to be linked to the operator. This would lead to 'natural' frames in the range of $A$, however they are poor for approximating $u^\dagger$. R. Steveson added to the requirements for choosing frames: the approximation properties have to be better than for each individual bases, computation must be computationally effective, frames must be sufficiently incoherent in order to allow for adaptive schemes (report on numerical tests using Schwarz iteration for overlapping grids (similar to old Randy Bank paper in the late 90's)). T. Raasch said that the 'communication' between the frame and the residuals is of great importance and requires the incoherence of cross-correlation between frame elements. O. Holtz pointed to connections to 'deterministic' compressed sensing approaches (conjecture of Bougain). P. Oswald asked how to design frames which are able of capturing both, transport and diffusion simultaneously in physical application).

## 4.3 Algorithms and Applications

*Michael Unser (EPFL – Lausanne, CH)*
*Bruno Torresani (Université de Provence – Marseille, FR)*



The development of efficient algorithms for large-scale sparse signal recovery is probably one of the greatest practical benefits of the intense research effort around the theme of compressed sensing. In addition to CS, these methods are relevant for a large variety of conventional imaging and inverse problems in general, often delivering state-of-the-art performance. Iterative soft-thresholding algorithms (ISTA) were introduced by image processing pioneers about a decade ago and have been improved systematically since then. An important step was to recognize that a good part of the required mathematical machinery for convex optimization (proximal operators, backward-forward splitting techniques) had been established in the 70s and was directly transposable to the present class of problems. We now have at our disposal a large number of variants of ISTA some of which provide orders of magnitude speed improvements. While the discussion group recognized the pressing need for a general purpose and highly-performant algorithm for $\ell_1$-type optimization, they also felt that a point of saturation had been reached and that it would difficult to come up with much faster schemes. What is also required is some careful benchmarking and critical comparison of methods. Yet, the participants also felt that there would still be space for creative engineering to come up with optimized schemes that take advantage of the specificities of certain classes of problems. More important than the algorithm is the problem formulation; in particular, the way of introducing prior information on the class of desirable solutions. It was pointed out that imposing sparsity analysis constraints would generally yield better results than the typical synthesis formulation (sparse generative model) that is central to CS. This is in line with the older concept of regularization as well as the Bayesian formulation of the reconstruction task. The panel did identify the following research challenges:

- the development of efficient large-scale optimization methods for extended classes of non-convex functionals
- the search for better sparsifying transforms and dictionaries (synthesis vs. analysis)
- the design of better regularization functionals
- (non-Gaussian) Bayesian formulations and the derivation/characterization of optimal estimators
- error and sensitivity analysis.

It was also pointed that it is in principle harder to design data-driven dictionaries for analysis purposes rather than synthesis. The concept of sparsity promotes simplicity; it provides a data-processing version of Occam's razor that is most attractive for algorithm design and probably here to stay (once the hype around CS has settled down). There are still many opportunities ahead for applying those techniques to real-world imaging and signal processing problems, beyond the typical tasks of denoising, deblurring and in-painting.

We are now essentially at the point where $\ell_1$-optimization is competitive speed-wise with the classical linear methods such as conjugate gradient.

## 5 Seminar schedule

| | | |
|---|---|---|
| Monday | Vetterli | Sampling in the Age of Sparsity |
| | Hansen | Generalized Sampling and Infinite Dimensional Compressed Sensing |
| | Unser | Stocastic models for sparse and piecewise smooth signals |
| | Pfander | Sensing, Local Approximation and Quantization of Operators with Bandlimited Kohn-Nirenberg Symbols |
| | Ehler | Random tight frames in directional statistics |
| | Pezeshki | Compressed Sensing for High Resolution Image Inversion |
| Tuesday | Oswald | Space Splittings and Schwarz-Southwell Iterations |
| | Raasch | Quarkonial frames of wavelet type - Stability and moment conditions |
| | Starck | Reconstructing and Analyzing Astrophysical Dark Matter Mass Maps using Sparsity |
| | Lim | Optimally sparse image approximations using compactly supported shearlet frames |
| | Zhuang | The Digital Shearlet Transform on Pseudo-polar Grids |
| | Lorenz | Basis pursuit denoising: Exact test instances and exact recovery for ill-posed problems |
| | | *Discussion groups* |
| Wednesday | Potts | Parameter estimation for exponential sums by approximate Prony method |
| | Kunis | Stable discretisations for sparse fast Fourier transforms |
| | Figueiredo | Alternating Direction Optimization for Imaging Inverse Problems with Sparsity-Inducing Regularization |
| Thursday | Plonka-Hoch | Sparse Approximation of Images by the Easy Path Wavelet Transform |
| | Bergmann | Multivariate Periodic Function Spaces |
| | Steidl | Halftoning, Quadrature Errors and Discrepancies |
| | Wojtaszczyk | How $\ell_1$ -minimisation for RIP matrices reacts to measurement errors? |
| | Nam | Cosparse Analysis Modeling |
| | Schiffler | Sparsity for ill-posed and ill-conditioned problems |
| | | *Summary of discussion groups* |
| Friday | Guleryuz | A Majorization-minimization algorithm for sparse factorization and some related applications |
| | Friedrich | Adaptive wavelet methods for inverse parabolic problems |

## Participants

- Ronny Bergmann
Universität zu Lübeck, DE
- Martin Ehler
National Institutes of Health - Bethesda, US
- Michael Elad
Technion - Haifa, IL
- Hans Georg Feichtinger
Universität Wien, AT
- Mario Figueiredo
TU Lisboa, PT
- Massimo Fornasier
RICAM - Linz, AT
- Ulrich Friedrich
Universität Marburg, DE
- Onur G. Guleryuz
DoCoMo USA Labs - Palo Alto, US
- Anders Hansen
University of Cambridge, GB
- Evelyn Herrholz
Hochschule Neubrandenburg, DE
- Olga Holtz
TU Berlin, DE
- Stefan Kunis
Universität Osnabrück, DE
- Gitta Kutyniok Universität Osnabrück, DE
- Wang-Q Lim
Universität Osnabrück, DE
- Dirk Lorenz
TU Braunschweig, DE
- Peter Maaß
Universität Bremen, DE
- Sangnam Nam
INRIA - Rennes, FR
- Peter Oswald
Jacobs University - Bremen, DE
- Ali Pezeshki
Colorado State University, US
- Götz E. Pfander
Jacobs University - Bremen, DE
- Gerlind Plonka-Hoch
Universität Göttingen, DE
- Daniel Potts
TU Chemnitz, DE
- Thorsten Raasch
Universität Mainz, DE
- Stefan Schiffler
Universität Bremen, DE
- Jean-Luc Starck
CEA - Gif sur Yvette, FR
- Gabriele Steidl
Universität Mannheim, DE
- Rob Stevenson
University of Amsterdam, NL
- Gerd Teschke
Hochschule Neubrandenburg, DE
- Bruno Torresani
Université de Provence, FR
- Michael Unser
EPFL - Lausanne, CH
- Martin Vetterli
EPFL - Lausanne, CH
- Przemyslaw Wojtaszczyk
University of Warsaw, PL
- Xiaosheng Zhuang
Universität Osnabrück, DE