AFRL-OSR-VA-TR-2013-0132

Building Mathematical Tools to Find Needles in Haystacks

Ingrid Daubechies, A. Robert Calderbank, Amit Singer

Princeton University

March 2013
Final Report

DISTRIBUTION A: Approved for public release.

AIR FORCE RESEARCH LABORATORY
AF OFFICE OF SCIENTIFIC RESEARCH (AFOSR)
ARLINGTON, VIRGINIA 22203
AIR FORCE MATERIEL COMMAND

REPORT DOCUMENTATION PAGE	*Form Approved* *OMB No. 0704-0188*

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
13-02-2013	FINAL	1 Sept 2009 - 31 Aug. 2012

4. TITLE AND SUBTITLE Building Mathematical Tools to Find Needles in Haystacks	**5a. CONTRACT NUMBER**
	5b. GRANT NUMBER FA9550-09-1-0551
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) Daubechies, Ingrid Calderbank, A. Robert Singer, Amit	**5d. PROJECT NUMBER**
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Princeton University Princeton, NJ 08540	**8. PERFORMING ORGANIZATION REPORT NUMBER**
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) AFOSR 875 N Randolph St Arlington, VA 22203	**10. SPONSOR/MONITOR'S ACRONYM(S)**
	11. SPONSOR/MONITOR'S REPORT NUMBER(S) AFRL-OSR-VA-TR-2013-0132

12. DISTRIBUTION/AVAILABILITY STATEMENT

DISTRIBUTION A: APPROVED FOR PUBLIC RELEASE

13. SUPPLEMENTARY NOTES

matrix completion, signal analysis, sensing, compressed sensing, sensor networks

14. ABSTRACT

The PIs and their students, postdocs and collaborators have worked on rigidity theory for sensor network localization, low rank matrix completion, sensor network localization in 3D and NMR, cryo electron microscopy, finite frame for signal processing, vector diffusion maps, and the detection of different "hands" in old master paintings. These are different frameworks in which large volumes of data have to be handled and analyzed; the work carried out by the PIs and their co-workers has resulted in the construction of and better insight in mathematical tools for these purposes.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON / 19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	U	SAR	Ingrid Daubechies / 9196602805

Reset

INSTRUCTIONS FOR COMPLETING SF 298

1. REPORT DATE. Full publication date, including day, month, if available. Must cite at least the year and be Year 2000 compliant, e.g. 30-06-1998; xx-06-1998; xx-xx-1998.

2. REPORT TYPE. State the type of report, such as final, technical, interim, memorandum, master's thesis, progress, quarterly, research, special, group study, etc.

3. DATES COVERED. Indicate the time during which the work was performed and the report was written, e.g., Jun 1997 - Jun 1998; 1-10 Jun 1996; May - Nov 1998; Nov 1998.

4. TITLE. Enter title and subtitle with volume number and part number, if applicable. On classified documents, enter the title classification in parentheses.

5a. CONTRACT NUMBER. Enter all contract numbers as they appear in the report, e.g. F33615-86-C-5169.

5b. GRANT NUMBER. Enter all grant numbers as they appear in the report, e.g. AFOSR-82-1234.

5c. PROGRAM ELEMENT NUMBER. Enter all program element numbers as they appear in the report, e.g. 61101A.

5d. PROJECT NUMBER. Enter all project numbers as they appear in the report, e.g. 1F665702D1257; ILIR.

5e. TASK NUMBER. Enter all task numbers as they appear in the report, e.g. 05; RF0330201; T4112.

5f. WORK UNIT NUMBER. Enter all work unit numbers as they appear in the report, e.g. 001; AFAPL30480105.

6. AUTHOR(S). Enter name(s) of person(s) responsible for writing the report, performing the research, or credited with the content of the report. The form of entry is the last name, first name, middle initial, and additional qualifiers separated by commas, e.g. Smith, Richard, J, Jr.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES). Self-explanatory.

8. PERFORMING ORGANIZATION REPORT NUMBER. Enter all unique alphanumeric report numbers assigned by the performing organization, e.g. BRL-1234; AFWL-TR-85-4017-Vol-21-PT-2.

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES). Enter the name and address of the organization(s) financially responsible for and monitoring the work.

10. SPONSOR/MONITOR'S ACRONYM(S). Enter, if available, e.g. BRL, ARDEC, NADC.

11. SPONSOR/MONITOR'S REPORT NUMBER(S). Enter report number as assigned by the sponsoring/ monitoring agency, if available, e.g. BRL-TR-829; -215.

12. DISTRIBUTION/AVAILABILITY STATEMENT. Use agency-mandated availability statements to indicate the public availability or distribution limitations of the report. If additional limitations/ restrictions or special markings are indicated, follow agency authorization procedures, e.g. RD/FRD, PROPIN, ITAR, etc. Include copyright information.

13. SUPPLEMENTARY NOTES. Enter information not included elsewhere such as: prepared in cooperation with; translation of; report supersedes; old edition number, etc.

14. ABSTRACT. A brief (approximately 200 words) factual summary of the most significant information.

15. SUBJECT TERMS. Key words or phrases identifying major concepts in the report.

16. SECURITY CLASSIFICATION. Enter security classification in accordance with security classification regulations, e.g. U, C, S, etc. If this form contains classified information, stamp classification level on the top and bottom of this page.

17. LIMITATION OF ABSTRACT. This block must be completed to assign a distribution limitation to the abstract. Enter UU (Unclassified Unlimited) or SAR (Same as Report). An entry in this block is necessary if the abstract is to be limited.

Final Report – AFOSR grant FA9550-09-1-0551

The PIs worked with their students and postdocs on several projects that received support from the AFOSR grant:

Deterministic Measurement Matrices for Compressed Sensing:
The aim of compressed sensing is to recover attributes of sparse signals using very few measurements. Given an overall bit budget for quantization, we have demonstrated that there is value to redundant measurement. With graduate students Sina Jafarpour and Victoria Kostina, and with postdoc Marco Duarte, we considered measurement matrices for which signal recovery is still possible even after dropping certain subsets of D measurements. We introduced the concept of a measurement matrix that is weakly democratic in the sense that the amount of information about the signal carried by each of the designated D-subsets is the same. We constructed examples of deterministic measurement matrices that are weakly democratic by exponentiating codewords from the binary second order Reed Muller code. The value in rejecting D measurements that are on average larger, is to be able to provide a finer grid for vector quantization of the remaining measurements, even after discounting the original budget by the bits used to identify the reject set. Simulation results demonstrate that redundancy improves recovery SNR, sometimes by a wide margin. Optimum performance occurs when a significant fraction of measurements are rejected. These findings were presented by Ms. Kostina at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011), Prague, Czech Republic in May 2011.

Sampling of Sparse Bandlimited Signals:
Technological constraints severely limit the rate at which analog-to-digital converters can reliably sample signals. Recently, Tropp et al. proposed an architecture called the Random Demodulator (RD), that attempts to overcome this obstacle for sparse bandlimited signals. Central to the RD architecture is a white noise-like, bipolar modulating waveform that changes polarity at a rate equal to the signal bandwidth. Since there is a hardware limitation to how fast analog waveforms can change polarity without undergoing shape distortion, this leads to the RD also having a constraint on the maximum allowable bandwidth. Together with graduate student Andrew Harms and postdoc Waheed Bajwa, we have proposed an extension of the RD, the constrained random demodulator (CRD) that bypasses this bottleneck by replacing the original modulating waveform with a run-length limited modulating waveform that changes polarity at a slower rate than the signal bandwidth. We prove that the CRD, despite employing a modulating waveform with correlations, enjoys theoretical guarantees that are quite similar to the original RD architecture. In addition, for a given sampling rate and rate of change in the modulating waveform polarity, numerical simulations confirm that the CRD can sample a signal with an even wider bandwidth without a significant loss in performance. These findings were presented by Mr. Harms at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011), Prague, Czech Republic in May 2011.

Finite Frames:
With graduate student Dustin Mixon and postdoc Waheed Bajwa, we investigated two

parameters that measure the coherence of a frame: worst-case and average coherence. We first provided a catalog of nearly tight frames with small worst-case and average coherence. Next, we found a new lower bound on worst-case coherence; we compare it to the Welch bound and used it to interpret recent signal reconstruction results. We then developed an algorithm that transforms frames in a way that decreases average coherence without changing the spectral norm or worst-case coherence. Finally, we used worst-case and average coherence to garner near-optimal probabilistic guarantees on both sparse signal detection and reconstruction in the presence of noise. These findings were presented by Mr. Mixon at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011), Prague, Czech Republic in May 2011.

Vector Diffusion Maps:

PI Singer and Hau-tieng Wu (graduated in Spring 2012, under the supervision of PI Daubechies) have developed Vector Diffusion Maps (VDM), a new mathematical framework for organizing and analyzing massive high dimensional data sets, images and shapes. VDM is a mathematical and algorithmic generalization of diffusion maps and other nonlinear dimensionality reduction methods, such as LLE, ISOMAP and Laplacian eigenmaps. While existing methods are either directly or indirectly related to the heat kernel for functions over the data, VDM is based on the heat kernel for vector fields. VDM provides tools for organizing complex data sets, embedding them in a low dimensional space, and interpolating and regressing vector fields over the data. In particular, it equips the data with a metric, referred to as the vector diffusion distance. In the manifold learning setup, where the data set is distributed on (or near) a low dimensional manifold, we prove the relation between VDM and the connection-Laplacian operator for vector fields over the manifold.

Sensor Network Localization in 3D and NMR Spectroscopy:

Calculations of structures from NMR data have used various approaches, most of which have significant limitations of precision or accuracy or speed, especially in calculating and evaluating ensembles of solution structures. PI Singer, together with graduate students Mihai Cucuringu and David Cowburn have applied a general application of the graph realization problem to this issue which we term 3D-ASAP (As-Synchronized-As-Possible). This is applied to the general three-dimensional problem of graph realization and the specific three-dimensional structure(s) of molecules from a collection of noisy short range experimental distances as a specific example. Numerical simulations show that 3D-ASAP compares favorably to other existing algorithms in terms of robustness to uncertainties in the measured distances, sparse connectivity of the distance matrix, and running times. These characteristics result from the non-incremental and non-iterative nature of 3D-ASAP. 3-D-ASAP starts by finding, embedding, and aligning uniquely localizable subsets of neighboring objects termed patches. The patches can be individual atoms or those groups of atoms that are rigidly related to each other. In the noise free case, each patch agrees with its global positioning up to an unknown range of motion of translation, rotation and possibly reflection. The reflections and rotations are estimated using our recently developed eigenvector synchronization algorithm, and translations are estimated by solving overdetermined linear system. The algorithm scales almost linearly as a number of atoms increases and can be implemented in a distributed fashion. 3D-

ASAP is able to exploit the additional information from rigid groups whose coordinates are previously known, for example, rigid peptide planes and ring systems. As a preliminary demonstration, we show its application to the reconstruction of 1D3Z ubiquitin PDB using such planar information. Further developments include the fitting of experimental observed NMR data and any other derived data items for modeling, and use in more complex ensemble calculations.

Cryo Electron Microscopy: Lanhui Wang continued working with PI Singer on three-dimensional structure determination of macromolecules from noisy cryo-electron microscopy projection images. More progress has been made this year, dealing with both simulated images and real microscope images provided to us by Sigworth Lab at Yale Medical School, specifically: 1) Linear and non-linear Wiener filtering of sinograms in the PCA domain leading to significantly improved detection rate of common-lines. 2) Detection of common lines between images from different defocus groups (different contrast functions). 3) Improved 3-D tomographic inversion algorithm, together with Cedric Vonesch (postdoctoral research associate working with PI Daubechies) and Yoel Shkolnisky (Tel Aviv University). The algorithm is based on two key components: a) a variational formulation that promotes sparsity in the wavelet domain and b) the Toeplitz structure of the combined projection/back-projection operator. The first idea has proven to be very effective for the recovery of piecewise-smooth signals, which is confirmed by our numerical experiments. The second idea allows for a computationally efficient implementation of the reconstruction procedure, using only one circulant convolution per iteration.

Detecting different "hands" in old master paintings: PI Daubechies used, with Marco Duarte and Josephine Wolff, algorithms that promote sparsity to identify style differences in the painted surface that correlate with different styles in the underdrawings in early 16th century Flemish paintings by Goswin van der Weyden. Art historians surmised that the different style underdrawings indicated whether or not he master himself (sketchy underdrawing) or an apprentice (very detailed underdrawing to prepare for surface painting) would finish the painting layer; an analysis of fine details in the brushwork of the top layer (without the underdrawing) did indeed allow a different classification, correlating strongly with the underdrawing style. Josephine Wolff's 2010 Senior Thesis in Mathematics, on Wavelet Analysis of Goswin van der Weyden Underdrawings, supervised by Calderbank and Daubechies, was awarded the Gregory T. Pope '80 Prize for science writing by the Princeton Council on Science and Technology. She presented a poster describing her research at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2011), Prague, Czech Republic in May 2011.

During the academic year, the PIs organized a weekly seminar that focused on high dimensional data analysis and applied harmonic analysis, including sparsity and compressive sampling. The seminar was regularly attended by PACM graduate students and postdoctoral research fellows.