Using Google Exacycle to Map a Universal Phase Transition in Compressive Sensing

1. Visitor's full name and contact information
Name: Sparseman
Contact Info: Just shine some structured light on some clouds. I'll make sense of it.
2. Visitor's affiliation (University, School, College and/or Department)
None.
3. Research abstract and goals

The goal of this undertaking is to provide the most comprehensive map of the Donoho-Tanner phase transition [3]. This map is key to understanding sparse solutions of underdetermined systems of linear equations and provide some guidance for compressive sensing sensor development.

Introduction

The new world of Big Data requires us to devise an ever increasing processing power in order to make sense of this constant « deluge ». Until very recently, the methods dedicated to reducing this sort of data did not keep up with the exponential increase in the size of said data. Starting in 2004 [1,2], a new field, compressive sensing (which includes sketching, random projections....), has allowed us to circumvent this bottleneck by introducing a well known assumption about our world directly in data acquisition methods and hardware [5]. This key assumption is that we live in a power law world where most sets are made up of few large elements and many tiny ones. In other words, the world can be seen as being "compressible". This insight and the work on compressive sensing merged into a powerful tool to investigate and understand our world by reducing the acquisition of data in the first place.

Within this context, Donoho and Tanner described a phase transition in [3] that puts limits on what is feasible with compressive sensing. Instead of paraphrasing their findings, here is an excerpt from [3]:

"...This curious phenomenon in high-dimensional geometric probability is one of a small number of fundamental such phase transitions. We claim they have consequences in several applied fields:

- *in selecting models for statistical data analysis of large datasets,*
- *in coping with outlying measurements in designed experiments,*
- *in determining how many samples we need to take in designing imaging*

devices.
The consequences can be both profound and important. They range from negative philosophical – if your database has too many 'junk' variables in it nothing can be learned from it – to positive-practical – it isn't really necessary to sit cooped up for an hour in a medical MRI scanner: with the right software, the necessary data could be collected in a fraction of the time commonly used today..."

While the example of MRI is helpful, it is just the tip of the iceberg [8]. A whole flurry of new sensors are being tried in different academic and industrial settings [5]. Some of these technologies will never make it to market, while others may be highly disruptive. The Donoho-Tanner phase transition provides a rule of thumb for the practicing engineer on how to go about these new sensors development.

Project description

We want to show that this phase transition boundary is universal in that it applies to a larger set of conditions than the ones found by Donoho and Tanner [3][4]. Our approach wil specifically consists in computing this phase transition with different:

- reconstruction solvers (L1/L0, Greedy, Iterative Hard Thresholding, ensemble approaches...)
- noise level (additive, folding, multiplicative)
- pre-modulation, measurement matrices and dictionaries [6]
- sparse vector distribution

Every computation performed for these phase transition boundaries do not rely on previous computations, hence the proposed work is "embarrassingly parallel". Through the Nuit Blanche blog [7] we expect specialists to help us define some of these boundaries and even play a more important role in these computations.

I. Description should include estimates of total resource as well as individual instance resource (CPU, RAM, I/O, data set) consumption

The study by Donoho and Tanner [3] used 6.8 CPU years but we want to look at a larger set of parameters (described above). Our experience is that current hardware featuring a 2.0GHz CPU and 2GB RAM are sufficient to perform computations of these phase transitions.
I/O requirements: Data In consists in the installation of a code(Octave script + Octave executable) on a target CPU and a very small text file pertinent to the specific parameters of interests. Data Out consists of a table / matrix less than 10 MB large.

II. Description of software required and evidence that it could be recompiled to Google Native Client

In order to quickly gear up our effort, we will use scripts written in the open source Octave software [9]. Since Octave is written in C++ and open source, we expect it to recompile the scripts on the Google Native Client flawlessly. However, some amount of time will be required to understand how to go from simple Octave scripts to fully compilable Google Native Clients.

III. Description of the data processing pipeline to prepare data for and process results from the CPU intensive part of Exacycle

Very little data needs to be prepared before the CPU intensive part of the Exacycle. Results produced from CPU intensive part of the Exacycle are eventually simply merged on one CPU where the phase transition can be assembled.

IV. Expected outcomes and results

The expected outcome of this effort will be to expand the domain in which the Donoho-Tanner phase transition is applicable and how it changes as a result of specific parameters such as noise level. These charts are likely to shape up the development of new sensor hardware (MRI, CT, hyperspectral imagery, acoustics,....)
Results of findings will be featured on Nuit Blanche [7], a blog with a small following of specialists of compressive sensing, and may potentially be publishable in ArXiv and eventually in the peer-review literature. All the data generated during the course of this study will be made freely available to the academic community.

References:

[1] E. J. Candes, J. Romberg, and T. Tao, “Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information,” IEEE Trans. Inf. Theory, vol. 52, pp. 489-509, 2006.
[2] D. L. Donoho, “Compressed sensing,” IEEE Trans. Inf. Theory, vol. 52, pp. 1289-1306, 2006
[3] Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing by Jared Tanner and David L. Donoho (http://www.maths.ed.ac.uk/~tanner/DoTa_Universality.pdf)
[4] Precise Undersampling Theorems, David L. Donoho and Jared Tanner.(http://ecos.maths.ed.ac.uk/papers/DoTa_PUT.pdf)
[5] A list of current Compressive Sensing hardware.
(http://sites.google.com/site/igorcarron2/compressedsensinghardware)
[6] Universal and Efficient Compressed Sensing Strategy through Spread Spectrum Modulation by Gilles Puy, Pierre Vandergheynst, Remi Gribonval, and Yves Wiaux.
(http://infoscience.epfl.ch/record/163850/files/manuscript_1.pdf?version=3)
[7] http://nuit-blanche.blogspot.com
[8] Fill in the Blanks: Using Math to Turn Lo-Res Datasets Into Hi-Res Samples by Jordan Ellenberg, WIRED Magazine, http://www.wired.com/magazine/2010/02/ff_algorithm/all/1
[9] Octave, Open source Software, http://www.octave.org