Abstract: Combinatorial Methods for Computationally Efficient Non-Linear Approximation

M. A. Iwen
Institute for Mathematics and its Applications (IMA)
University of Minnesota
iwen@ima.umn.edu

May 21, 2009

We consider combinatorial methods for the efficient generation of near-optimal signal approximations using a minimal number of linear measurements. Related applications include deterministic sublinear-time approximate Fourier algorithms [6], noise-tolerant sketching techniques for massive data sets [9], signal denoising/compression [8], and compressed sensing techniques [2, 3] for imaging with radically reduced sensor requirements [5]. Related theoretical results include generalized theorems concerning the complexity of matrix multiplication [7].

Consider a signal, $\vec{x} \in \mathbb{C}^N$, which is compactly represented in some known orthonormal basis $\mathcal{B}$ with associated $N \times N$ unitary matrix Ψ. We can rank the basis elements, $\vec{b}_j \in \mathcal{B}$, from most to least able to represent $\vec{x}$ by

$$\left|\langle \vec{b}_1, \vec{x} \rangle\right| \geq \left|\langle \vec{b}_2, \vec{x} \rangle\right| \geq \cdots \geq \left|\langle \vec{b}_N, \vec{x} \rangle\right|. \tag{1}$$

If we wish to approximate $\vec{x}$ using only $k \ll N$ basis elements from $\mathcal{B}$ we will obtain optimal results by using $\vec{b}_1, \dots, \vec{b}_k$ above and approximating $\vec{x}$ by

$$\vec{x}_{\mathcal{B},k} = \sum_{j=1}^{k} \langle \vec{b}_j, \vec{x} \rangle \cdot \vec{b}_j. \tag{2}$$

Here $\vec{x}_{\mathcal{B},k}$ is an optimal k-term non-linear approximation to $\vec{x}$ in terms of $\mathcal{B}$. Note that finding $\vec{x}_{\mathcal{B},k}$ is equivalent to calculating the k-largest magnitude entries of $\Psi\vec{x}$.

Given a basis, $\mathcal{B}$, for $\mathbb{C}^N$ and an positive integer $k \ll N$ we will discuss combinatorial constructions for rectangular $m \times N$ measurement matrices with $m \ll N$ rows. Given such a measurement matrix, $\mathbf{M}^k_{\mathcal{B}}$, we will see that we can accurately encode any signal, $\vec{x} \in \mathbb{C}^N$, as a compressed m-length vector

$$\vec{y} = \mathbf{M}^k_{\mathcal{B}} \cdot \vec{x}.$$

A related fast algorithm then can take $\vec{y}$ as input and output an accurate $O(k)$-term approximation to $\vec{x}_{\mathcal{B},k}$. Note that minimizing the size of m is critical not only as a means of reducing measurement acquisition/storage costs, but also as a means of decreasing the approximation algorithm's runtime. Typically m is $(k \cdot \log N)^{O(1)}$, where the constant factors involved depend on the desired approximation error, the presence/absence of additional knowledge concerning the input, etc.. Fast combinatorial algorithms for approximating $\vec{x}_{\mathcal{B},k}$ typically run in $O(m \cdot \log N)^{O(1)}$-time. Hence, such fast combinatorial non-linear approximation algorithms are *sublinear-time* with respect to the ambient signal dimension N.

1 Proposed Talk Outline

As above, we will begin by introducing and motivating combinatorial methods for the non-linear approximation of signals. We will then discuss non-linear approximation algorithms motivated by combinatorial group testing [4] techniques. These ideas will ultimately lead us to the following useful definition.

Definition 1. *A collection, $\mathcal{S}$, of subsets of $[0, N) \cap \mathbb{N}$ is called k-**majority selective** if for all $X \subset [0, N) \cap \mathbb{N}$ with $|X| \leq k$ and all $n \in [0, N) \cap \mathbb{N}$, more than half of the subsets $S \in \mathcal{S}$ containing n are such that $S \cap X = \{n\} \cap X$ (i.e., every $n \in [0, N) \cap \mathbb{N}$ is separated from all (other) members of X in more than half of the $\mathcal{S}$ subsets containing n).*

As we shall see, it turns out that k-majority selective collections of subsets provide noise tolerant methods for approximating signals, $\vec{x} \in \mathbb{C}^N$, that exhibit polynomial compressibility in a known basis $\mathcal{B}$ (i.e., that have $\left|\langle \vec{b}_j, \vec{x} \rangle\right| \leq C \cdot j^{-p}$ for some $C \in \mathbb{R}^+$ and $p > 1$ in Equation 1 above). However, the resulting measurement matrices have a number of rows, m, that depends on $\frac{1}{p-1}$. As a result, methods based on k-majority selective collections of subsets can require a large number of measurements to discover useful information regarding signals which are insufficiently compact under $\mathcal{B}$ (e.g., are compressible with $p \approx 1$).

In the final portion of our talk we will discuss more recent combinatorial methods based on unbalanced expander graphs (e.g., see [1]). These methods are capable of quickly gathering useful signal information for signals of arbitrary compactness (e.g., compressible with $p \approx 1$) using a bounded number of measurements $m = (k \cdot \log N)^{O(1)}$. The following theorem provides an example of the types of approximation guarantees enjoyed by recent unbalanced expander based methods.

Theorem 1. ***(Berinde, Gilbert, Indyk, Karloff, Strauss [1])*** *Suppose $\mathcal{B}$ is a basis for $\mathbb{R}^N$. Given $\vec{x} \in \mathbb{R}^N$, unbalanced expander based algorithms may be used to approximate $\vec{x}_{\mathcal{B},k}$ (see Equation 2) by a $O(k/\epsilon)$-sparse output signal, $\vec{a}_{\mathcal{B},k,\epsilon}$, which satisfies*

$$||\vec{x} - \vec{a}_{\mathcal{B},k,\epsilon}||_1 \leq (1 + 3\epsilon)||\vec{x} - \vec{x}_{\mathcal{B},k}||_1.$$

These algorithms take input $\mathbf{M}^k_{\mathcal{B}} \cdot \vec{x}$, where $\mathbf{M}^k_{\mathcal{B}}$ is an $\left(\frac{k}{\epsilon} \cdot \log N\right)^{O(1)} \times N$ matrix, and run in $\left(\frac{k}{\epsilon} \cdot \log N\right)^{O(1)}$-time.

Fix ϵ, and let Ψ be the $N \times N$ unitary matrix associated with the basis $\mathcal{B}$ from Theorem 1. In effect, Theorem 1 guarantees that an unbalanced expander based algorithm will locate $O(k)$ of the dominant entries in $\Psi \cdot \vec{x}$. If $\Psi \cdot \mathbf{A}$ contains k significant entries whose values collectively dominate all others combined, then these most significant entries will be found and their values will be well approximated. If $\Psi \cdot \vec{x}$ has no dominant set of k entries, then Theorem 1 only guarantees a $O(k)$-sparse representation will be returned which is trivially bounded. However, in such cases sparse non-linear approximation is generally a hopeless task anyways, and a bounded, albeit poor, sparse representation is the best one can expect.

References

[1] R. Berinde, A. C. Gilbert, P. Indyk, H. Karloff, and M. Strauss. Combining geometry and combinatorics: A unified approach to sparse signal recovery. *Submitted*, 2008.

[2] E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52:489–509, 2006.

[3] D. Donoho. Compressed Sensing. *IEEE Trans. on Information Theory*, 52:1289–1306, 2006.

[4] D. Z. Du and F. K. Hwang. *Combinatorial Group Testing and Its Applications*. World Scientific, 1993.

[5] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 83, March 2008.

[6] M. A. Iwen. A deterministic sub-linear time sparse fourier algorithm via non-adaptive compressed sensing methods. In *Proc. of ACM-SIAM symposium on Discrete algorithms (SODA'08)*, 2008.

[7] M. A. Iwen and C. Spencer. A note on compressed sensing and the complexity of matrix multiplication. *Information Processing Letters*, 109, 2009.

[8] S. Mallet. *A Wavelet Tour of Signal Processing*. China Machine Press, 2003.

[9] S. Muthukrishnan. Data Streams: Algorithms and Applications. *Foundations and Trends in Theoretical Computer Science*, 1, 2005.