# Statement of Research

Eric Price

December 14, 2012

My goal in research is to discover theoretical insights that can guide practitioners in the creation of useful systems. To this end, I try to focus on relatively simple algorithms that are feasible to implement and have small big-Oh constants; when finding lower bounds, I look for ones that give guidance in the creation of efficient algorithms. To calibrate my understanding of the relation between theory and practice, I implement about half my algorithms and analyze their empirical performance.

With this in mind, I have decided to focus my research on the interdisciplinary area of *sparse recovery*, which includes aspects of compressive sensing and streaming algorithms. The goal of sparse recovery is to acquire and process "sparse" data from a small number of samples. The topic offers the opportunity to develop theoretical and mathematical techniques that apply to big data problems in diverse areas such as data stream analysis and signal processing. On the theory side, sparse recovery involves fascinating techniques from algorithms, statistics, probability theory, and information theory. On the application side, sparse recovery lets me interact with researchers that understand the practical constraints involved in creating real systems. It is a great area to develop theoretical results that can make an impact in practice.

My first paper in the area shows fundamental limitations on the "standard" compressive sensing framework, by showing that the number of samples required by the seminal work of Candès, Romberg, and Tao is in fact *optimal*. This led me to investigate two methods for circumventing these lower bounds: making the sampling process *adaptive* and incorporating additional structural assumptions on the signals. I have shown that adaptivity enables a significant—and in some cases *exponential*—reduction in the number of samples required for sparse recovery. I have also shown the first *linear time* algorithm to exploit one of the most common additional structural assumptions. My research on algorithms that are highly efficient in both number of samples and processing time has culminated in algorithms to compute the *Fourier transform* efficiently when its output is sparse. These algorithms are faster than the ubiquitous Fast Fourier Transform for moderately sparse data, both in theory and in practice.

## 1 Sparse recovery

The amount of data we attempt to collect, store, and process grows exponentially over time, leading to severe resource bottlenecks. Fortunately, many real world signals are "sparse," meaning that most of the information is concentrated in relatively few coordinates of the signal. Sparsity has long been used by compression algorithms to decrease storage costs; the goal of *sparse recovery* is to use it to decrease data collection and processing costs. Sparse recovery allows signal acquisition with fewer measurements than naive observation, at the cost of requiring the ability to take *linear measurements* $Ax$ of the signal vector $x$. This observation model is powerful enough to significantly improve measurement efficiency, but simple enough to be implemented in diverse settings. In particular, it yields space-efficient algorithms for processing massive data streams and time-efficient algorithms for computing Fourier transforms.

Sparse recovery can be applied in almost any setting where linear measurements are feasible and the vectors of interest are sparse. For example, in *streaming algorithms* the matrix product $Ax$ can be maintained under increments and decrements to $x$, and sparse recovery solves the fundamental "heavy hitters" problem [CCF02, CM06, GI10]. In *camera design*, linear measurements correspond to (for example) placing a filter in front of the lens, and are implementable in prototype hardware like the single pixel camera [DDT+08]. In *medical imaging*, the nuclear magnetic resonance used in MRI inherently gives Fourier samples of the image [LDSP08]. In *genetic testing*, blood samples can be mixed together prior to testing to observe linear combinations of the binary vector corresponding to genes [ECG+09].

Formally, sparse recovery considers the acquisition of signals that are approximately *sparse* in some well-chosen basis (a vector $x \in \mathbb{R}^n$ is $k$-*sparse* if $x$ has at most $k$ non-zero coefficients, and approximately $k$-sparse if it is close to a $k$-sparse vector). Measurements are allowed to be arbitrary linear combinations of coefficients of $x$. The algorithm must recover an approximation $x^*$ to $x$ with error proportional to the error in a $k$-sparse approximation of $x$. The mathematical formulation of the problem is to choose a matrix $A \in \mathbb{R}^{m \times n}$ such that, for any vector $x \in \mathbb{R}^n$, using the $m \ll n$ observations $Ax$ one can recover $x^*$ satisfying

$$\|x^* - x\| \leq C \min_{k\text{-sparse } x_k} \|x_k - x\| \tag{1}$$

for some approximation factor $C = \Theta(1)$ and distance metric $\|\cdot\|$ that is usually $\ell_1$ or $\ell_2$; if $x$ is exactly $k$-sparse, this will recover it exactly, and if $x$ is approximately $k$-sparse, this will recover it approximately. We also allow *randomized* constructions, where $A$ is chosen from some distribution and recovery must succeed with "good" probability over the choice of $A$.

## 2 Results

**Lower bounds.** The main goal of sparse recovery is to minimize the number of measurements. The seminal work of Candès, Romberg, and Tao [CRT06] presented a method with $O(k \log(n/k))$ measurements. Over the next few years, a number of results improved the generality of the matrices or the speed of reconstruction, but none managed to decrease the number of measurements. My first work in the field showed that this dependence is *optimal* [DIPW10] for the standard $\ell_1$ and $\ell_2$ metrics. Note that in the deterministic setting, the lower bound was previously known [Kaš77, GG84, BDDW08]; we extended it to the general case of randomized constructions of $A$. We later developed new upper and lower bounds to tighten the dependence on the approximation factor $C$ in this standard setting [PW11].

The latter work featured a particularly simple lower bound technique based on the communication capacity of the Gaussian channel. This technique has proven to be quite general, leading to lower bounds for moment estimation [PW12], adaptive measurements [PW13], and adaptive Fourier measurements [HIKP12a].

**Adaptivity.** By the above, any improvement on the number of measurements requires changing the model to circumvent the lower bound. A natural modification is to allow *adaptive* measurements, where each row of the matrix may be chosen based on the results of previous rows. Adaptivity is feasible in many sparse recovery contexts (for example, it corresponds to multiple pass streaming algorithms), so previous researchers had studied it and found empirical benefits [MSW08, AWZ08] and an improved dependence on the approximation factor $C$ [HBCN09]. Using intuition from our lower bound [PW11], we created an algorithm featuring the *first general asymptotic improvement using adaptivity*, achieving $O(k \log\log(n/k))$ measurements [IPW11]. The improvement is not just asymptotic: in our experiments with $k = 1$, adaptivity already gives a factor 2 improvement at $n = 8192$. Furthermore, we showed that our result is *optimal* in this $k = 1$ case [PW13].

**Model-based compressive sensing.** Another way to improve on the number of measurements is to assume the input vector $x$ is not only sparse but also comes from a "nice" distribution. One important example is *block sparsity*, where the support of $x$ lies in $k/\log n$ contiguous "blocks" of length $\log n$. This models "bursty" signals, where the "activity" is localized in a small number of blocks. In this setting, $O(k)$ measurements suffice [BCDH10]. We created an algorithm for this task with running time $O(n)$, improving on the previous $\widetilde{O}(nk)$ [Pri11][1].

Another model of structure satisfied by many real signals is *power law decay*, a.k.a. Zipf's law. In this model, the $i$th largest coefficient has magnitude proportional to $i^{-c}$ for constant $c > 1/2$. In many settings, sparsity is merely a proxy for power laws [BCDH10]. We showed that for such vectors, the classic Count-Sketch [CCF02] algorithm used in practice (e.g. at Google [PDGQ05]) gives a stronger guarantee than was previously known [MP12]. While the standard analysis of Count-Sketch uses a simple variance bound to control the *maximum* error in the estimation of any coordinate, we used Fourier analysis to show that the *average* error is asymptotically less than the maximum. This yields substantial improvements in the overall estimation error when the signal has power law decay, improving our theoretical understanding of the performance of an algorithm used in practice.

**Earth-Mover's distance.** Beyond optimizing the number of measurements, another goal is to extend the guarantee of sparse recovery. The usual distance metric for the sparse recovery guarantee is $\ell_1$ or $\ell_2$. When the signal is an image, however, the $\ell_p$ metrics can be poor approximations of visual similarity. To address this concern in image retrieval, some researchers use a different metric known as *Earth Mover's Distance* [RTG00]. We constructed an algorithm that efficiently performs sparse recovery with respect to Earth Mover's Distance using $O(k \log(n/k))$ measurements [IP11].

While interesting in its own right, this work also has interesting connections to other areas. It turns out that the problem of sparse recovery under Earth Mover's Distance is nearly equivalent to that of finding a "coreset" for the $k$-medians problem. This latter problem is a generalization of $k$-median clustering, which has been well studied in the streaming model [FS05, Ind04]. Our result improves on the best known in this area.

**Sparse Fourier transform.** The Fast Fourier Transform (FFT) is a ubiquitous computational tool that computes the Discrete Fourier Transform of a given vector $y \in \mathbb{C}^n$ in $O(n \log n)$ time. It is widely used in many applications, including signal processing and compression (e.g., in the JPEG and MPEG standards). The reason for its use in compression is that for signals $y$ of interest, the discrete Fourier transform $\widehat{y}$ is concentrated among a small number $k$ of coordinates. A natural question that arises is: can we speed up image/video compression by estimating the $k$ large coordinates of $\widehat{y}$ in time closer to $k$ than to $n$? This problem is essentially equivalent to sparse recovery with measurements restricted to be rows of the (inverse) Fourier matrix.

A number of previous papers have studied sparse Fourier transforms, with the best achieving roughly $O(k \log^4 n)$ time [GMS05]. This is only faster than the $O(n \log n)$ fast Fourier transform when $k/n < O(1/\log^3 n)$; such extreme sparsity is rarely the case in the applications motivating this research.

Using insights derived from streaming algorithms and compressive sensing, we gave an algorithm to compute sparse Fourier transforms in $O(k \log(n/k) \log n)$ time [HIKP12a]. This is faster than the FFT for any sparsity ratio $k/n$ less than some fixed constant. In the special case where the signal is *exactly* $k$-sparse (i.e, has only $k$ non-zero Fourier terms), we gave a variant with the essentially optimal running time of $O(k \log n)$. Our implementation of the latter is faster than the fastest implementation of the FFT for $k/n < 3\%$ [HIKP12b]; not quite the $k/n \approx 7\%$ typical in image compression, but getting close.

[1]We use $\widetilde{O}(f)$ to denote $O(f \log^c f)$ for some constant $c$.

## 3 Research Agenda

Sparse recovery is a quickly developing area spanning computer science, statistics, and signal processing. The cross-fertilization of ideas between these fields makes it rich with open problems having many applications. Here are some of the topics I am particularly interested in:

**Fourier measurement complexity.** In medical imaging applications such as MRI, minimizing the sample complexity of the sparse Fourier transform is more important than minimizing the time complexity. During MRIs, the patient often needs to lie still for 30 minutes while Fourier samples of the image are observed. Thus improving the measurement complexity of MRIs would directly improve patient comfort and MRI cost. The sparse Fourier transform is an idealized version of the problem. My work has shown that $O(k \log(n/k) \log n)$ samples are sufficient [HIKP12a], and it is also known that with $O(k \log n)$ samples we can achieve an approximation factor $C = \Theta(\log^2 n)$ [CP11]. It remains open whether $O(k \log n)$ samples are sufficient to get a constant factor approximation.

For the two dimensional Fourier transform—which is more applicable to imaging applications—the gap is larger, because the natural extension of [HIKP12a] requires $O(k \log(n/k) \log^2 n)$ samples. To reduce this gap, we have shown an algorithm for two dimensional Fourier transforms using the optimal $O(k \log n)$ samples on *random inputs* at $k = \Theta(\sqrt{n})$ [GHI$^+$12]. I would like to get a similar result for worst case inputs and general $k$.

**Fast, generic algorithms.** So far, my work has focused on very specific measurement matrix designs tightly coupled with recovery algorithms. This is the norm in the computer science literature on sparse recovery because streaming algorithms allow full control of the sensing matrix. However, the other applications of sparse recovery involve physical hardware, so diverse real world constraints arise on the sensing matrix. For this reason, the statistics and signal processing communities have worked on generic recovery algorithms that apply to general classes of matrices, most often the set of matrices satisfying the *Restricted Isometry Property* or RIP. These general algorithms are slower than the matrix-specific designs and only support deterministic matrices, not random distributions on matrices. Featuring deterministic matrices is in some ways a benefit, but makes it impossible to achieve goals such as $\ell_2$ error bounds or improvements via adaptivity. I would like to combine the best features of both approaches.

One line of research in this direction is that of *fast RIP matrices*, which allow generic recovery algorithms to finish more quickly. One can show that both subsampled Fourier [CT06, RV08, CGV13] and subsampled circulant [KMR12] matrices satisfy the RIP and allow sparse recovery in only $\widetilde{O}(n)$ time rather than the naive $\widetilde{O}(nk)$. The drawback of this approach is that the number of measurements increases from the optimal $O(k \log(n/k))$ to $O(k \log^4 n)$ in both cases. My coauthors and I have narrowed the gap, giving a construction of RIP matrices featuring $\widetilde{O}(n)$ multiplication time and only $O(k \log^3 n)$ rows [NPW12]; by a black box reduction [KW11], our result also improves on the best known *fast Johnson-Lindenstrauss matrices* [AC09, AL11, KW11] used in dimensionality reduction. I want to close the remaining $\log^2 n$ gap, which seems to come from the gap between Dudley's entropy integral and Talagrand's generic chaining in probability theory.

**Perceptual coding.** Image compression standards such as JPEG or JPEG2000 are effective not just because they exploit the structure of images, but also because the errors they introduce are designed to be less perceptible to the human eye. Our work on Earth Mover's Distance gave hints at how sparse recovery could incorporate the same idea, both by making the norm more perceptually relevant and by introducing errors in a similar way to the JPEG2000 image format. Left open are both direct questions, such as developing tight upper and lower bounds for this metric, and general questions, such as developing a more comprehensive theory of perceptual sparse recovery.

**Tight constants.** The theory I have described is asymptotic in nature and ignores constants. But constants matter: for many signals, the improvement from the naive $n$ measurements to sparse recovery's $k \log(n/k)$ measurements is only one order of magnitude and can be swallowed up by sloppy constants. A similar issue is faced in coding theory, and just as we can compute channel capacities, we ought to be able to find the exact constant in sparse recovery. Our information theory-based lower bound techniques, being similar to coding theory ones, can likely get tight constants. Our analysis of algorithms, on the other hand, tends to use techniques that require sloppy constants. One encouraging result is [BJ10], which gives a sparse recovery algorithm with a tight constant that works on a restricted class of signals. This algorithm is used to create a code for the Gaussian channel, strengthening the connection between sparse recovery and coding and giving hope that tight constants might be achievable in general.

## References


[AC09] N. Ailon and B. Chazelle. The fast Johnson–Lindenstrauss transform and approximate nearest neighbors. *SIAM J. Comput.*, 39(1):302–322, 2009.

[AL11] N. Ailon and E. Liberty. Almost optimal unrestricted fast Johnson-Lindenstrauss transform. In *SODA*, pages 185–191, 2011.

[AWZ08] A. Aldroubi, H. Wang, and K. Zarringhalam. Sequential adaptive compressed sampling via Huffman codes. *Preprint*, 2008.

[BCDH10] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde. Model-based compressive sensing. *IEEE Transactions on Information Theory*, 56, No. 4:1982–2001, 2010.

[BDDW08] R. G. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, 28:253–263, 2008.

[BJ10] A. R. Barron and A. Joseph. Sparse superposition codes: Fast and reliable at rates approaching capacity with Gaussian noise. *Technical Report, available at* `http://www.stat.yale.edu/~aj268/Research.html`, November 2010.

[CCF02] M. Charikar, K. Chen, and M. Farach-Colton. Finding frequent items in data streams. In *Proceedings of the 29th International Colloquium on Automata, Languages and Programming (ICALP)*, pages 693–703, 2002.

[CGV13] M. Cheraghchi, V. Guruswami, and A. Velingker. Restricted isometry of Fourier matrices and list decodability of random linear codes. In *SODA (to appear)*, 2013. full version at CoRR abs/1207.1140.

[CM06] G. Cormode and S. Muthukrishnan. Combinatorial algorithms for Compressed Sensing. In *Proc. 40th Ann. Conf. Information Sciences and Systems*, Princeton, Mar. 2006.

[CP11] E.J. Candès and Y. Plan. A probabilistic and RIPless theory of compressed sensing. *Information Theory, IEEE Transactions on*, 57(11):7235–7254, 2011.

[CRT06] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8), 2006.

[CT06] E. J. Candès and T. Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52:5406–5425, 2006.

[DDT+08] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. G. Baraniuk. Single-pixel imaging via compressive sampling. *IEEE Signal Processing Magazine*, 2008.

[DIPW10] K. Do Ba, P. Indyk, E. Price, and D. Woodruff. Lower bounds for sparse recovery. *SODA*, 2010.

[ECG+09] Y. Erlich, K. Chang, A. Gordon, R. Ronen, O. Navon, M. Rooks, and G. J. Hannon. DNA Sudoku—harnessing high-throughput sequencing for multiplexed specimen analysis. *Genome Research*, 2009.

[FS05] G. Frahling and C. Sohler. Coresets in dynamic geometric data streams. *STOC*, 2005.

[GG84] A. Y. Garnaev and E. D. Gluskin. On the widths of the Euclidean ball. *Soviet Mathematics Doklady*, 30:200–203, 1984.

[GHI+12] B. Ghazi, H. Hassanieh, P. Indyk, D. Katabi, E. Price, and L. Shi. Sample-optimal average-case sparse fourier transform in two dimensions. *Unpublished manuscript*, 2012.

[GI10] A. Gilbert and P. Indyk. Sparse recovery using sparse matrices. *Proceedings of IEEE*, 2010.

[GMS05] A. Gilbert, M. Muthukrishnan, and M. Strauss. Improved time bounds for near-optimal space Fourier representations. *SPIE Conference, Wavelets*, 2005.

[HBCN09] J. Haupt, R. G. Baraniuk, R. Castro, and R. Nowak. Compressive distilled sensing. *Asilomar*, 2009.

[HIKP12a] H. Hassanieh, P. Indyk, D. Katabi, and E. Price. Nearly optimal sparse fourier transform. *STOC*, 2012.

[HIKP12b] H. Hassanieh, P. Indyk, D. Katabi, and E. Price. sFFT: Sparse Fast Fourier Transform. *http://groups.csail.mit.edu/netmit/sFFT/*, 2012.

[Ind04] P. Indyk. Algorithms for dynamic geometric problems over data streams. *STOC*, 2004.

[IP11] P. Indyk and E. Price. K-median clustering, model-based compressive sensing, and sparse recovery for earth mover distance. *STOC*, 2011.

[IPW11] P. Indyk, E. Price, and D. Woodruff. On the power of adaptivity in sparse recovery. *FOCS*, 2011.

[Kaš77] B. S. Kašin. The widths of certain finite-dimensional sets and classes of smooth functions. *Izv. Akad. Nauk SSSR Ser. Mat.*, 41(2):334–351, 478, 1977.

[KMR12] F. Krahmer, S. Mendelson, and H. Rauhut. Suprema of chaos processes and the restricted isometry property. *arXiv*, abs/1207.0235, 2012.

[KW11] F. Krahmer and R. Ward. New and improved Johnson-Lindenstrauss embeddings via the Restricted Isometry Property. *SIAM J. Math. Anal.*, 43(3):1269–1281, 2011.

[LDSP08] M. Lustig, D.L. Donoho, J.M. Santos, and J.M. Pauly. Compressed sensing mri. *Signal Processing Magazine, IEEE*, 25(2):72–82, 2008.

[MP12] G. T. Minton and E. Price. Improved concentration bounds for count-sketch. *arXiv preprint arXiv:1207.5200*, 2012.

[MSW08] D. M. Malioutov, S. Sanghavi, and A. S. Willsky. Compressed sensing with sequential observations. *ICASSP*, 2008.

[NPW12] J. Nelson, E. Price, and M. Wootters. New constructions of RIP matrices with fast multiplication and fewer rows. *arXiv preprint arXiv:1211.0986*, 2012.

[PDGQ05] R. Pike, S. Dorward, R. Griesemer, and S. Quinlan. Interpreting the data: Parallel analysis with sawzall. *Scientific Programming*, 13(4):277, 2005.

[Pri11] E. Price. Efficient sketches for the set query problem. *SODA*, 2011.

[PW11] E. Price and D. Woodruff. (1+eps)-approximate sparse recovery. *FOCS*, 2011.

[PW12] E. Price and D. Woodruff. Applications of the Shannon-Hartley theorem to data streams and sparse recovery. In *Information Theory Proceedings (ISIT), 2012 IEEE International Symposium on*, pages 2446–2450. IEEE, 2012.

[PW13] E. Price and D. Woodruff. Lower bounds for adaptive sparse recovery. *SODA (to appear)*, 2013.

[RTG00] Y. Rubner, C. Tomasi, and L. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, 2000.

[RV08] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Communications on Pure and Applied Mathematics*, 61(8):1025–1045, 2008.