Sparse Recovery for Earth Mover Distance

Rishi Gupta
CSAIL MIT

Piotr Indyk
CSAIL MIT

Eric Price
CSAIL MIT

December 5, 2010

Abstract

We initiate the study of sparse recovery problems under the *Earth-Mover Distance (EMD)*. Specifically, we design a distribution over $m \times n$ matrices A, for $m \ll n$, such that for any x, given Ax, we can recover a k-sparse approximation to x under the EMD distance. We also provide an empirical evaluation of the method that, in some scenarios, shows its advantages over the "usual" recovery in the ℓ_p norms.

1 Introduction

In recent years, a new "linear" approach for obtaining a succinct approximate representation of n-dimensional vectors (or signals) has been discovered. For any signal x, the representation is equal to Ax, where A is an $m \times n$ matrix, or possibly a random variable chosen from some distribution over such matrices. The vector Ax is often referred to as the *measurement vector* or *linear sketch* of x. Although m is typically much smaller than n, the sketch Ax often contains plenty of useful information about the signal x.

A particularly useful and well-studied problem is that of *stable sparse recovery*. The problem is typically defined as follows: for some norm parameters p and q and an approximation factor $C > 0$, given Ax, recover a vector $\hat{x}$ such that

$$\|x - \hat{x}\|_p \le C \cdot \mathrm{Err}_k^q(x) \tag{1}$$

where $\mathrm{Err}_k^q(x) = \min_{k\text{-sparse } x'} \|x - x'\|_q$. Note that we say that x' is k-sparse if it has at most k non-zero coordinates. Sparse recovery has applications to numerous areas such as data stream computing [Mut03, Ind07] and compressed sensing [CRT06, Don06], notably for constructing imaging systems that acquire images directly in compressed form (e.g., [DDT$^+$08, Rom09]). The problem has been a subject of extensive study over the last few years, with the goal of designing schemes that enjoy good "compression rate" (i.e., low values of m) as well as good algorithmic properties (i.e., low encoding and recovery times). It is known by now[1] that there exist matrices A and associated recovery algorithms that produce approximations $\hat{x}$ satisfying Equation (1) with $\ell_p = \ell_q = \ell_1$, constant approximation factor C and sketch length $m = O(k \log(n/k))$; it is also known that this sketch length is asymptotically optimal [DIPW10, FPRU10]. Results for other combinations of ℓ_p/ℓ_q norms are known as well.

However, limiting the error measures to variants of ℓ_p norms is quite inconvenient in many applications. First, the distances induced by ℓ_p norms are typically only quite raw approximations of the perceptual differences between images. As a result, in the field of computer vision, several more elaborate notions have been proposed (e.g., in [RTG00, Low04, Lyu05, GD05]). Second, there are natural classes of images for which the distances induced by the ℓ_p norm are virtually meaningless. For example, consider images of "point clouds", e.g., obtained via astronomical imaging. If we are given two such images, where each point in the second image is obtained via small random translation of a point in the first image, then the ℓ_p distance between the images will be close to the largest possible, even though the images are quite similar to each other.

Motivated by the above considerations, we initiate the study of sparse recovery under non-ℓ_p distances. In particular, we focus on the *Earth-Mover Distance* (EMD) [RTG00]. Informally, for the case of two-dimensional $\Delta \times \Delta$ images (say, $x, y : [\Delta]^2 \to \mathbb{R}^+$) which have the same ℓ_1 norm, the EMD is defined as the cost of the min-cost flow that transforms x into y, where the cost of transporting a "unit" of mass from pixel $p \in [\Delta]^2$ of x to a pixel $q \in [\Delta]^2$ of y is equal to the ℓ_2 distance[2] between p and q. Earth-

[1] In particular, a random Gaussian matrix [CRT06] or a random sparse binary matrix ([BGI$^+$08], building on [CCFC02, CM04, CM06]) has this property with overwhelming probability. See [GI10] for an overview.

[2] One can also use the ℓ_1 distance. Note that the two distances differ by at most a factor of $\sqrt{2}$ for two-dimensional images.

Mover Distance and its derivatives are popular metrics for comparing similarity between images, feature sets, etc. [RTG00, GD05].

We define sparse recovery under EMD in a way similar to sparse recovery under ℓ_p norms (as per Eq. 1). Specifically, the goal is to construct a distribution over $m \times n$ matrices A, $n = \Delta^2$, such that for any vector x, given Ax, one can reconstruct a vector $\hat{x}$ such that

$$\mathrm{EMD}(x, \hat{x}) \leq C \cdot \mathrm{Err}_k^{\mathrm{EMD}}(x)$$

with constant probability, where $\mathrm{Err}_k^{\mathrm{EMD}}(x) = \min_{k\text{-sparse } x'} \mathrm{EMD}(x, x')$.

Discussion and connections. What does sparse recovery with respect to the EMD distance mean? As it turns out, the task has the following natural interpretation. Let x' be the minimizer of $\mathrm{EMD}(x, x')$ over all k-sparse vectors. Then one can observe that the non-zero entries of x' correspond to the cluster centers in the best k-median[3] clustering of x. Moreover, for each such center c, the value of x'_c is equal to the total weight of pixels in the cluster centered at c. Thus, a solution to the k-median problem can provide a solution to our sparse recovery problem as well.

There has been prior work on the k-median problem in the streaming model under insertions and deletions of points [FS05, Ind04]. Such algorithms utilize linear sketches, and therefore implicitly provide schemes for approximating the k-medians of x from a linear sketch of x (although they do not necessarily provide the cluster weights, which are needed for the sparse recovery problem). Both algorithms yield a method for approximating the k-medians from $\Omega(k^2)$ measurements. Our result gives a constant approximation to the k-medians problem without the quadratic dependence on k.

2 Algorithms

On a high level, our approach is to reduce the sparse recovery problem under EMD to sparse recovery under ℓ_1. The reduction is performed by using a "pyramid" mapping P [IT03, GD05] (building on [Cha02]), which provides a multi-resolution representation of the image. The mapping P is defined as follows. First we impose $\log \Delta + 1$ nested grids G_i on $[\Delta^2]$. For each $i = 0 \ldots s$, $s = \log \Delta$, the grid G_i is a partition of the image into cells of side length 2^i. For each i, we define a mapping P_i such that each entry in $P_i x$ corresponds to a cell c in G_i, and its value is equal to the sum of coordinates of x falling into c. The final mapping P is defined as

$$Px = [2^0 P_0 x, 2^1 P_1 x, \ldots, 2^s P_s x]$$

It is known [Cha02, IT03] that (a randomized version of) the mapping P has the property that $\|Px - Py\|_1$ approximates $\mathrm{EMD}(x, y)$ up to some (super-constant) factor. In our case we need only a weaker property, namely that for any x there is an $O(ks)$-sparse z such that $\|Px - z\|_1 \leq D \mathrm{Err}_k^{\mathrm{EMD}}(x)$. Building on [FS05] we can show that this property holds for a constant D. It follows that we can find a k-sparse approximation to x under EMD by finding an $O(ks)$-sparse approximation to Px under ℓ_1. This can be achieved using roughly $O(ks \log \Delta) = O(k \log^2 \Delta)$ measurements using known techniques.

However, thanks to the properties of the mapping P, there is space for improvement. First, if the vector x is non-negative, so is Px; this constrains the space of potential solutions. Second, we can exploit the particular form of vectors Px. Specifically, the coordinates of Px have the following hierarchical structure: the coordinates are organized into a full r-ary tree for $r = 2^2$, and each internal node is *twice* the sum of its children. Moreover, if x is k-sparse, then the non-zero coefficients of Px are connected in a tree-like fashion. This enables us to constrain the sparsity patterns of the recovered approximation to Px, in a manner similar to *model-based compressive sensing* [BCDH10].

By incorporating these constraints into the sparse recovery algorithms, we significantly reduce the number of necessary measurements. Specifically, our algorithm augments the SSMP algorithm from [BI09] in the following manner. The SSMP algorithm proceeds by performing incremental updates to its current solution (say, z), interleaved with periodic sparsification of z, i.e., zeroing out all but its top k entries. Our modified version follows the same outline. However during the sparsification process, it imposes the following constraints (i) the coefficients should be non-negative and (ii) the non-zero entries should have a tree structure. That is, the sparsification process greedily computes a connected tree of $O(k)$ coefficients, and zeros out all other coefficients.

3 Empirical evaluation

We performed preliminary experiments investigating the proposed approach[4]. For the data vectors x

[3] For completeness, in our context the k-median is defined as follows. First, each pixel $p \in [n]^2$ is interpreted as a point with weight x_p. Then the goal is to find a set $C \subset [n]^2$ of k "medians" that minimizes the objective function $\sum_{p \in [n]^2} \min_{c \in C} \|p - c\|_2 x_p$.

[4] Source code for the experiments is available from http://web.mit.edu/ecprice/www/papers/allerton10/

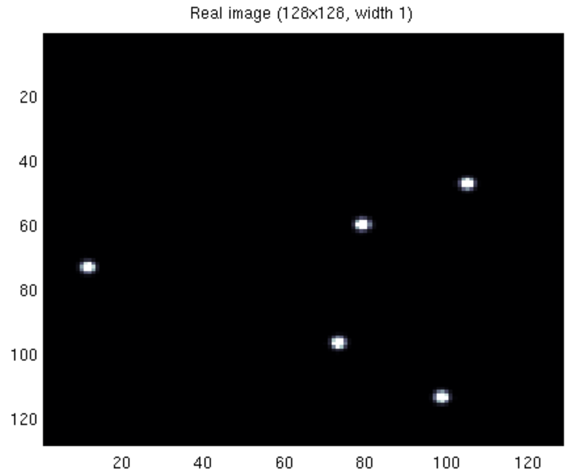

Figure 1: Example of a synthetic image used in our experiments. The image resolution is 128 x128, and the number of clusters k is equal to 5. The standard deviation of the clusters is equal to 1.

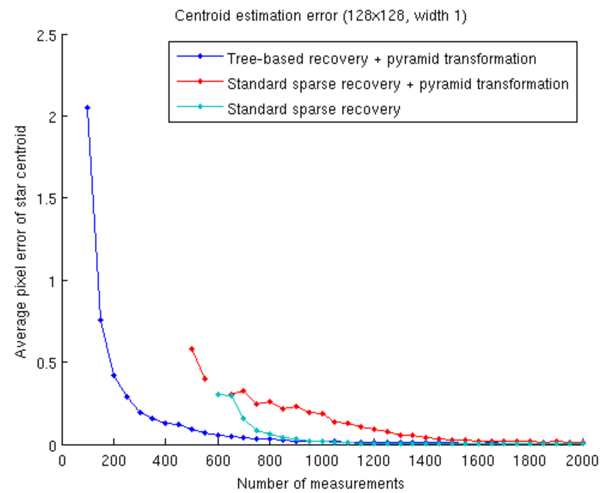

Figure 2: Cluster center estimation error for different recovery methods, as a function of the number of measurements m. For each m we perform 15 runs of the respective recovery algorithm, and plot the median (over the runs) of the average distance from each recovered center to the closest actual one. For low values of m most runs resulted in an insufficient number of recovered centers. In such cases the distance is assumed to be infinite and not displayed.

we used synthetic "star-like" images, generated from mixtures of Gaussians placed in random positions of an image, with a prespecified width (i.e., standard deviation). An example image is given in Figure 1.

The results are presented in Figure 2. For three different algorithms, it shows the cluster center estimation error as a function of the number of measurements m. The three algorithms are as follows (i) sparse recovery applied to Px, incorporating the aforementioned positivity and tree-sparsity constraints; (ii) "standard" sparse recovery applied to Px; and (iii) standard sparse recovery on x. We used SSMP [BI09] as a basis for all three algorithms.

One can observe that the algorithm (i) provides non-trivial results for much lower values of m than the other two algorithms. Moreover, when all three algorithms can be used, the algorithm (i) typically yields much lower estimation error.

Acknowledgements

The authors would like to thank Yaron Rachlin, Tye Brady and Ben Lane from Draper Lab for numerous conversations and help with the data.

This research has been supported in part by a David and Lucille Packard Fellowship, MADALGO (the Center for Massive Data Algorithmics, funded by the Danish National Research Association) and NSF grant CCF-0728645. R. Gupta has been supported in part by a Draper Fellowship. E. Price has been supported in part by a Cisco Fellowship and NSF Graduate Research Fellowship.

References

[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.

[BGI+08] R. Berinde, A. Gilbert, P. Indyk, H. Karloff, and M. Strauss. Combining geometry and combinatorics: a unified approach to sparse signal recovery. *Allerton*, 2008.

[BI09] R. Berinde and P. Indyk. Sequential sparse matching pursuit. *Allerton*, 2009.

[CCFC02] M. Charikar, K. Chen, and M. Farach-Colton. Finding frequent items in data streams. *ICALP*, 2002.

[Cha02] M. Charikar. Similarity estimation techniques from rounding. In *STOC*, pages 380–388, 2002.

[CM04] G. Cormode and S. Muthukrishnan. Improved data stream summaries: The count-min sketch and its applications. *Latin*, 2004.

[CM06] G. Cormode and S. Muthukrishnan. Combinatorial algorithms for compressed sensing. *Sirocco*, 2006.

[CRT06] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1208–1223, 2006.

[DDT+08] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. 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.

[Don06] D. L. Donoho. Compressed Sensing. *IEEE Trans. Info. Theory*, 52(4):1289–1306, Apr. 2006.

[FPRU10] S. Foucart, A. Pajor, H. Rauhut, and T. Ullrich. The gelfand widths of lp-balls for $0 < p \leq 1$. *preprint*, 2010.

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

[GD05] K. Grauman and T. Darrell. The pyramid match kernel: Discriminative classification with sets of image features. *ICCV*, 2005.

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

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

[Ind07] P. Indyk. Sketching, streaming and sublinear-space algorithms. *Graduate course notes, available at* `http://stellar.mit.edu/S/course/6/fa07/6.895/`, 2007.

[IT03] P. Indyk and N. Thaper. Fast color image retrieval via embeddings. *Workshop on Statistical and Computational Theories of Vision (at ICCV)*, 2003.

[Low04] D. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[Lyu05] S. Lyu. Mercel kernels for object recognition with local features. *CVPR*, 2005.

[Mut03] S. Muthukrishnan. Data streams: Algorithms and applications (invited talk at soda'03). *Available at http://athos.rutgers.edu/∼muthu/stream-1-1.ps*, 2003.

[Rom09] J. Romberg. Compressive sampling by random convolution. *SIAM Journal on Imaging Science*, 2009.

[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.