OUTLIER IDENTIFICATION VIA RANDOMIZED ADAPTIVE COMPRESSIVE SAMPLING

Xingguo Li and Jarvis D. Haupt

Department of Electrical and Computer Engineering
University of Minnesota, Twin Cities
Minneapolis, Minnesota USA 55455
{lixx1661, jdhaupt}@umn.edu

ABSTRACT

This paper examines the problem of locating outlier columns in a large, otherwise low-rank, matrix. We propose a simple two-step adaptive sensing and inference approach and establish theoretical guarantees for its performance. Our results show that accurate outlier identification is achievable using very few linear summaries of the original data matrix – as few as the squared rank of the low-rank component plus the number of outliers, times constant and logarithmic factors. We demonstrate the performance of our approach experimentally in two stylized applications, one motivated by robust collaborative filtering tasks, and the other by saliency map estimation tasks arising in computer vision and automated surveillance.

Index Terms— Adaptive and compressive sensing, robust PCA

1. INTRODUCTION

In this paper we address a matrix *outlier identification* problem. Suppose $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ is a data matrix that may be expressed as

$$\mathbf{M} = \mathbf{L} + \mathbf{C}, \tag{1}$$

where $\mathbf{L}$ is a low-rank matrix, and $\mathbf{C}$ is a matrix of outliers that is nonzero in only a fraction of its columns. We are ultimately interested in identifying the locations of the nonzero columns of $\mathbf{C}$ from a small number of linear measurements of $\mathbf{M}$, with a particular focus on settings where $\mathbf{M}$ may be very large.

Our investigation is motivated in part by robust collaborative filtering applications, in which the goal may be to identify the locations (or even quantify the number) of corrupted data points or outliers in a large data array. Such tasks may arise in a number of contemporary applications, for example, when identifying malicious responses in survey data or anomalous patterns in network traffic, to name a few. Depending on the nature of the outliers, conventional low-rank approximation approaches based on principal component analysis (PCA) [1] may be viable options for these tasks, but they become increasingly computationally demanding as the data become very high-dimensional. Here, our aim is to leverage dimensionality reduction ideas along the lines of those used in randomized numerical linear algebra, (e.g., [2, 3]) and compressed sensing (e.g., [4, 5]), in order to reduce the size of the data on which our approach operates.

We are also motivated by an image processing task that arises in computer vision and surveillance applications – that of identifying the "saliency map" of a given image [6–9]. In contrast to existing methods designed to identify saliency map as a "post processing" step, our aim here is to estimate the saliency map *directly from compressive samples*. We address this problem using a linear subspace-based saliency model, where images are represented as matrices whose columns are *vectorized* versions of image patches (see, e.g., [10]). The efficacy of modeling salient regions as outliers from a single common low-dimensional subspace has been established recently in [11]. Our approach may find utility in rapid threat detection in surveillance applications where the goal is to identify regions of anomalous behavior rather to image the entire scene.

This work was supported by the NSF under Award No. CCF-1217751.

1.1. Innovations and Our Approach

We propose an approach that employs dimensionality reduction techniques within the context of a two-step adaptive sampling and inference procedure, and our method is based on a few key insights. First, we exploit the fact that the enabling geometry of our problem (to be defined) is approximately preserved if we operate on a "compressed" version $\mathbf{\Phi M}$ of $\mathbf{M}$ that has potentially many fewer rows. Next, we use the fact that we can learn the linear subspace spanned by the columns of the low rank component of $\mathbf{\Phi M}$ using a small subset of the columns of $\mathbf{\Phi M}$. Our algorithmic approach for this step utilizes a recently proposed method called *Outlier Pursuit* (OP) [12] that aims to separate a given matrix, say $\mathbf{Y}$, into its low-rank and column-sparse components using the convex optimization

$$\underset{\boldsymbol{L},\boldsymbol{C}}{\operatorname{argmin}} \ \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{C}\|_{1,2} \ \text{ s.t. } \mathbf{Y} = \boldsymbol{L} + \boldsymbol{C} \tag{2}$$

where $\|\boldsymbol{L}\|_*$ denotes the nuclear norm of $\boldsymbol{L}$, $\|\boldsymbol{C}\|_{1,2}$ is the sum of the ℓ_2 norms of the columns of $\boldsymbol{C}$, and $\lambda > 0$ is a regularization parameter. Finally, we leverage the fact that we may effectively transform – via linear observations – the overall outlier identification problem into a compressed sensing problem, and employ existing theoretical results (e.g., [13]) to establish the overall success of our approach.

In more detail, our procedure is initialized by selecting (ultimately, random) matrices $\mathbf{\Phi} \in \mathbb{R}^{m \times n_1}$ with $m < n_1$, $\mathbf{A} \in \mathbb{R}^{p \times n_2}$ with $p < n_2$ and vector $\boldsymbol{\phi} \in \mathbb{R}^{1 \times m}$, and by generating a column subsampling matrix $\mathbf{S}$ comprised of a (random) subset of columns of the $n_2 \times n_2$ identity matrix. In a first step, we observe $\mathbf{\Phi M S}$ and apply the *Outlier Pursuit* procedure to this observed data in an attempt to recover the linear subspace spanned by columns of $\mathbf{\Phi L}$ (denoted here by $\widehat{\mathcal{L}}_{(1)}$). Then, in a second step, we use a linear measurement operator whose net effect is to *(i)* reduce the overall $n_1 \times n_2$ matrix $\mathbf{M}$ to a $1 \times n_2$ vector $\boldsymbol{\phi}\, \mathbf{P}_{\widehat{\mathcal{L}}_{(1)}^{\perp}}\, \mathbf{\Phi M}$ where $\mathbf{P}_{\widehat{\mathcal{L}}_{(1)}^{\perp}}$ denotes projection onto the orthogonal complement of $\widehat{\mathcal{L}}_{(1)}$, and *(ii)* compressively sample the resulting (row) vector (post-multiplying with $\mathbf{A}^T$). We then apply sparse inference methods to recover the locations of the nonzeros of $\boldsymbol{\phi}\, \mathbf{P}_{\widehat{\mathcal{L}}_{(1)}^{\perp}}\, \mathbf{\Phi M}$, which nominally correspond to locations of the outlier columns of $\mathbf{M}$. We call our approach Adaptive Compressive Outlier Sensing (ACOS), and provide a detailed description as Algorithm 1. Our main contributions are a theoretical analysis of ACOS and experimental evaluation of its performance.

Algorithm 1 Adaptive Compressive Outlier Sensing (ACOS)

Input: $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$, column sampling Bernoulli parameter $\gamma \in [0,1]$, regularization parameter $\lambda > 0$, measurement matrices $\mathbf{\Phi} \in \mathbb{R}^{m \times n_1}$, $\mathbf{A} \in \mathbb{R}^{p \times n_2}$, measurement vector $\boldsymbol{\phi} \in \mathbb{R}^{1 \times m}$

Initalize: Column sampling matrix $\mathbf{S} = \mathbf{I}_{:,\mathcal{S}}$, where $\mathcal{S} = \{i : S_i = 1\}$ with $\{S_i\}_{i \in [n_2]}$ i.i.d. Bernoulli(γ)

Step 1

Collect Measurements: $\mathbf{Y}_{(1)} = \mathbf{\Phi M S}$

Solve: $\{\widehat{\mathbf{L}}_{(1)}, \widehat{\mathbf{C}}_{(1)}\} = \underset{\boldsymbol{L},\boldsymbol{C}}{\operatorname{argmin}} \|\boldsymbol{L}\|_* + \lambda \|\boldsymbol{C}\|_{1,2}$ s.t. $\mathbf{Y}_{(1)} = \boldsymbol{L} + \boldsymbol{C}$

Let: $\widehat{\mathcal{L}}_{(1)}$ be the linear subspace spanned by col's of $\widehat{\mathbf{L}}_{(1)}$

Step 2

Compute: $\mathbf{P}_{\widehat{\mathcal{L}}_{(1)}}$, the orthogonal projector onto $\widehat{\mathcal{L}}_{(1)}$

Set: $\mathbf{P}_{\widehat{\mathcal{L}}_{(1)}^{\perp}} \triangleq \mathbf{I} - \mathbf{P}_{\widehat{\mathcal{L}}_{(1)}}$

Collect Measurements: $\mathbf{y}_{(2)} = \boldsymbol{\phi}\, \mathbf{P}_{\widehat{\mathcal{L}}_{(1)}^{\perp}} \mathbf{\Phi M A}^T$

Solve: $\widehat{\mathbf{c}} = \underset{\mathbf{c}}{\operatorname{argmin}} \ \|\mathbf{c}\|_1$ s.t. $\mathbf{y}_{(2)} = \mathbf{c}\mathbf{A}^T$

Output: $\widehat{\mathcal{I}}_{\mathbf{C}} = \{i : \widehat{\mathrm{c}}_i \neq 0\}$

1.2. Relation to Prior Work

Our effort here leverages results from Compressive Sensing (CS), where the *sparsity* in the signal being acquired is exploited to devise efficient procedures for reconstructing high-dimensional objects [4, 5, 13]. The adaptive nature of our proposed approach is inspired by numerous recent works in the burgeoning area of adaptive sensing (see, for example, [14–23]). Our efforts here utilize a generalization of the notion of sparsity, formalized in terms of a low-rank plus outlier matrix model. In this sense, our efforts here are related to earlier work in Robust PCA [24, 25] that seek to identify low-rank matrices in the presence of sparse impulsive outliers, and their extensions where the outliers present as entire columns of an otherwise low-rank matrix [12, 26–28]. In fact, the computational approach and theoretical analysis of the first step of our approach make direct utilization of the results of [12].

1.3. Outline

The remainder of the paper is organized as follows. In Section 2 we formalize our problem and state our main theoretical result. Section 3 contains an experimental evaluation of our approach on synthetic data, as well as in a stylized image processing application of saliency map estimation. In Section 4 we provide discussion of a few potential future directions.

2. MAIN RESULTS

Our specific problem of interest here may be formalized as follows. Suppose $\mathbf{M} \in \mathbb{R}^{n_1 \times n_2}$ admits a decomposition of the form $\mathbf{M} = \mathbf{L} + \mathbf{C}$. Here, $\mathbf{L}$ is a matrix having rank at most r, and $n_{\mathbf{L}} \leq n_2$ nonzero columns. $\mathbf{C}$ is a column-sparse matrix with $k \leq n_2$ nonzero "outlier" columns that may occur only at the set of locations where the corresponding column of $\mathbf{L}$ is zero. Clearly, $n_{\mathbf{L}} \leq n_2 - k$. The notion that nonzero columns of $\mathbf{C}$ be "outliers" is codified as follows. Let $\mathcal{L}$ denote the linear subspace of $\mathbb{R}^{n_1}$ spanned by the columns of $\mathbf{L}$, and $\mathcal{L}^{\perp}$ denote its orthogonal complement in $\mathbb{R}^{n_1}$. Let $\mathbf{P}_{\mathcal{L}}$ and $\mathbf{P}_{\mathcal{L}^{\perp}}$ be the orthogonal projection operators onto $\mathcal{L}$ and $\mathcal{L}^{\perp}$, respectively. We assume that the nonzero columns of $\mathbf{C}$ occur at the indices $i \in \mathcal{I}_{\mathbf{C}} \triangleq \{i : \|\mathbf{P}_{\mathcal{L}^{\perp}} \mathbf{C}_{:,i}\|_2 > 0\}$, where $\mathbf{C}_{:,i}$ denotes the i-th column of $\mathbf{C}$. With this setup, our problem of interest here may be stated concisely – our aim is to identify the set $\mathcal{I}_{\mathbf{C}}$.

2.1. Assumptions

It is well-known in the matrix completion and robust PCA literature that separation of low-rank and sparse matrices from observations of their sum may not be a well-posed task – for example, matrices having only a single nonzero element are simultaneously low rank and sparse. To overcome these identifiability issues, it is common to assume that the linear subspace spanned by the rows and/or columns of the low-rank matrix be "incoherent" with the canonical basis (see, e.g., [12, 24–26, 29], among others). Here, we adopt a similar approach, and assume such a condition on the row space of the low-rank component $\mathbf{L}$ via the following definition from [12].

Definition 2.1 (Column Incoherence Property)**.** *Let $\mathbf{L} \in \mathbb{R}^{n_1 \times n_2}$ be a rank r matrix with at most $n_{\mathbf{L}} \leq n_2$ nonzero columns, and compact singular value decomposition (SVD) $\mathbf{L} = \mathbf{U\Sigma V}^*$, where $\mathbf{U}$ is $n_1 \times r$, $\mathbf{\Sigma}$ is $r \times r$, and $\mathbf{V}$ is $n_2 \times r$. The matrix $\mathbf{L}$ is said to satisfy the* column incoherence property *with parameter $\mu_{\mathbf{L}}$ if*

$$\max_i \|\mathbf{V}^* \mathbf{e}_i\|_2^2 \leq \mu_{\mathbf{L}} \frac{r}{n_{\mathbf{L}}}, \tag{3}$$

where $\{\mathbf{e}_i\}$ are basis vectors of the canonical basis for $\mathbb{R}^{n_2}$.

Note that $\mu_{\mathbf{L}} \in [1, n_{\mathbf{L}}/r]$; the lower limit is achieved when all elements of $\mathbf{V}^*$ have the same amplitude, and the upper limit when any one element of $\mathbf{V}^*$ is equal to 1 (i.e., when the row space of $\mathbf{L}$ is aligned with the canonical basis).

With this, we may state our structural assumptions concisely, as follows: we assume that the components $\mathbf{L}$ and $\mathbf{C}$ of the matrix $\mathbf{M} = \mathbf{L} + \mathbf{C}$ satisfy the following *structural conditions*:

- (**c1**) $\operatorname{rank}(\mathbf{L}) = r$,
- (**c2**) $\mathbf{L}$ has $n_{\mathbf{L}}$ nonzero columns,
- (**c3**) $\mathbf{L}$ satisfies the *column incoherence property* with parameter $\mu_{\mathbf{L}}$, and
- (**c4**) $|\mathcal{I}_{\mathbf{C}}| = k$.

2.2. Recovery Guarantees and Implications

Our main result identifies conditions under which the procedure outlined in Algorithm 1 succeeds. Our particular focus is on measurement matrices satisfying the following property.

Definition 2.2 (Distributional Johnson-Lindenstrauss (JL) Property)**.** *An $m \times n$ matrix $\mathbf{\Phi}$ is said to satisfy the* distributional JL property *if for any fixed $\mathbf{v} \in \mathbb{R}^n$ and any $\epsilon \in (0,1)$,*

$$\Pr\left(\left| \|\mathbf{\Phi v}\|_2^2 - \|\mathbf{v}\|_2^2 \right| \geq \epsilon \|\mathbf{v}\|_2^2 \right) \leq 2e^{-mf(\epsilon)}, \tag{4}$$

where $f(\epsilon) > 0$ is a constant depending only on ϵ that is specific to the distribution of $\mathbf{\Phi}$.

Certain random matrices are well-known to satisfy the distributional JL property and, as noted in [30], for many randomly constructed $\mathbf{\Phi}$, (e.g., such that entries of $\mathbf{\Phi}$ are i.i.d. zero-mean sub-Gaussian distribution), $f(\epsilon)$ is quadratic in ϵ as $\epsilon \to 0$. This general framework also allows us to directly utilize other specially constructed *fast* or *sparse* JL transforms [31, 32].

With this, we are in position to formulate our main result. We state the result here as a theorem; its proof appears in [33].

Theorem 2.1. *Suppose* $\mathbf{M} = \mathbf{L} + \mathbf{C}$*, where* $\mathbf{L}$ *and* $\mathbf{C}$ *satisfy the structural conditions (**c1**)-(**c4**) with* $k \leq \frac{1}{40(1+121\ r\mu_{\mathbf{L}})}\ n_2$*. For any* $\delta \in (0,1)$*, if the column subsampling parameter* γ *satisfies*

$$\gamma \geq \max\left\{\frac{1}{20},\ \frac{200\log(\frac{5}{\delta})}{n_{\mathbf{L}}},\ \frac{24\log(\frac{10}{\delta})}{n_2},\ \frac{10r\mu_{\mathbf{L}}\log(\frac{5r}{\delta})}{n_{\mathbf{L}}}\right\}, \tag{5}$$

the measurement matrices $\mathbf{\Phi}$ *and* $\mathbf{A}$ *are each drawn from any distribution satisfying* (4) *with* m *and* p *satisfying*

$$m \geq \frac{5(r+1)+\log(k)+\log(2/\delta)}{f(1/2)}, \tag{6}$$

$$p \geq \frac{11k+2k\log(n_2/k)+\log(2/\delta)}{f(1/4)}, \tag{7}$$

the elements of ϕ *are i.i.d. realizations of any continuous random variable, and for any upper bound* $k_{\rm ub}$ *of* k *the regularization parameter is set to* $\lambda = \frac{3}{7\sqrt{k_{\rm ub}}}$*, then the following hold simultaneously with probability at least* $1-3\delta$*:*

- *the ACOS procedure in Algorithm 1 correctly identifies the salient columns of* $\boldsymbol{C}$ *(i.e.,* $\widehat{\mathcal{I}}_{\boldsymbol{C}} = \mathcal{I}_{\boldsymbol{C}}$*), and*
- *the total number of measurements collected is no greater than* $\left(\frac{3}{2}\right)\gamma m n_2 + p$*.*

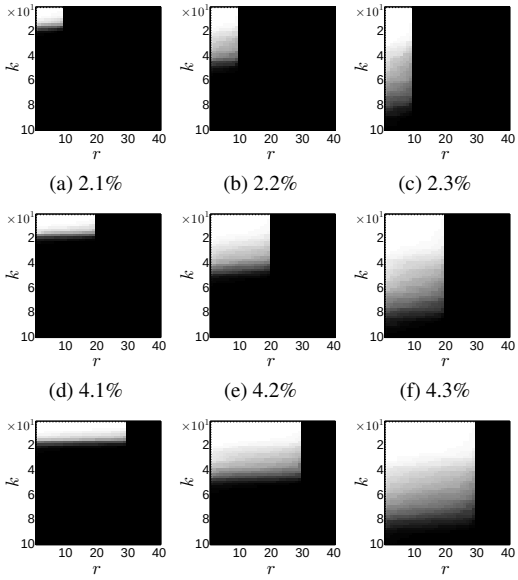
Our main result establishes that the ACOS approach succeeds with high probability with an effective sampling rate of $\frac{\#\text{obs}}{n_1 n_2} = \mathcal{O}\left(\max\left\{\frac{(r+\log k)(n_2/n_{\mathbf{L}})\mu_{\mathbf{L}} r\log r}{n_1 n_2}, \frac{(r+\log k)}{n_1}\right\} + \frac{k\log(n_2/k)}{n_1 n_2}\right)$, which may be small when r and k are each small relative to the problem dimensions. Compared with related work [26], which identifies both the column space and the set of outlier columns of a matrix $\mathbf{M} = \mathbf{L} + \mathbf{C}$ from $\mathcal{O}\left(n_2 r^2 \mu^2 \log(n_2)\right)$ samples of $\mathbf{M}$, we see that the sufficient conditions for the sample complexity of our approach are smaller than that of [26] by a factor of at least $1/r$, and, our approach does not require the row incoherence assumption. We provide some additional, experimental, comparisons between our ACOS method and the RMC method in Section 3.

3. EXPERIMENTAL EVALUATION

In this section we provide an experimental evaluation of the performance of our approaches for both synthetically generated and real data. We compare our methods with the Outlier Pursuit (OP) approach of [12] and the Robust Matrix Completion (RMC) approach of [26]. We implement the RMC method using an accelerated approximate alternating direction method of multipliers (ADMM) method inspired by [34] (as well as [12, 35]). We implement the OP methods (and intermediate executions of the OP-like optimization in Step 1 of our approach) using the procedure in [26]. We implement the ℓ_1-regularized estimation in Step 2 of our procedure as a LASSO problem solved using an accelerated proximal gradient method [35].

3.1. Synthetic Data

We experiment on synthetically generated $n_1 \times n_2$ matrices $\mathbf{M}$, with $n_1 = 100$ and $n_2 = 1000$, formed as follows. For a specified rank r and number of outliers k, we let the number of nonzero columns of $\mathbf{L}$ be $n_{\mathbf{L}} = n_2 - k$, generate two random matrices $\mathbf{U} \in \mathbb{R}^{n_1 \times r}$ and $\mathbf{V} \in \mathbb{R}^{n_{\mathbf{L}} \times r}$ with i.i.d. $\mathcal{N}(0,1)$ entries, and we take $\mathbf{L} = [\mathbf{U}\mathbf{V}^T\ \mathbf{0}_{n_1 \times k}]$. We generate the outlier matrix $\mathbf{C}$ as $\mathbf{C} = [\mathbf{0}_{n_1 \times n_{\mathbf{L}}}\ \mathbf{W}]$ where $\mathbf{W} \in \mathbb{R}^{n_1 \times k}$ has i.i.d. $\mathcal{N}(0,r)$ entries (which are also independent of entries of $\mathbf{U}$ and $\mathbf{V}$). Then, we set $\mathbf{M} = \mathbf{L} + \mathbf{C}$. Notice that the outlier vector elements have been scaled, so that all columns of $\mathbf{M}$ have the same squared ℓ_2 norm, in expectation. In all experiments we generate ϕ, $\mathbf{\Phi}$, and $\mathbf{A}$ with i.i.d. zero-mean Gaussian entries.

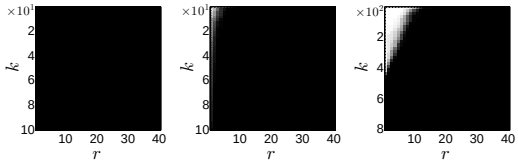


(a) 2.1% (b) 2.2% (c) 2.3%

(d) 4.1% (e) 4.2% (f) 4.3%

(g) 6.1% (h) 6.2% (i) 6.3%

Fig. 1. Outlier recovery phase transitions plots for ACOS (white regions correspond to successful recovery).



(a) 5% (RMC) (b) 10% (RMC) (c) 20% (RMC)

Fig. 2. Outlier recovery phase transitions plots for RMC.

Our first experiment investigates the "phase transition" behavior of our approach. The results in Figure 1 depict the outcome of this experiment for the 9 different sampling regimes examined, with experimental setting as follows. First, we fix the column downsampling fraction $\gamma = 0.2$, and choose a row sampling parameter $m \in \{0.1n_1, 0.2n_1, 0.3n_1\}$ (from top to bottom) and column sampling parameter $p \in \{0.1n_2, 0.2n_2, 0.3n_2\}$ (from left to right). Then, for each (r,k) pair with $r \in \{1,2,3,\ldots,40\}$ and $k \in \{2,4,6,\ldots,100\}$ we generate $\mathbf{M}$ as above, and for each of 3 different values of $\lambda \in \{0.3, 0.4, 0.5\}$ we perform 100 trials of Algorithm 1 recording in each whether the recovery approach succeeded[1] in identifying the locations of the true outliers for that value of λ. Then, at each (r,k) point examined we identify the point-wise maximum of the average success rates for the 3 different values of λ to assess the recovery achievability for *some* choice of regularization parameters. For easy comparison, we provide the average sampling rate as fraction of observations obtained (relative to the full matrix dimension) in the caption of each subfigure.

[1] We solve the optimization associated with Step 2 as a LASSO problem, with 10 different regularization parameters $\mu \in (0,1)$. We deem any trial a success if for at least one value of μ, there exists a threshold $\tau > 0$ such that $\min_{i \in \mathcal{I}_{\mathbf{C}}} |\widehat{\mathbf{c}}_i(\mu)| > \tau > \max_{j \notin \mathcal{I}_{\mathbf{C}}} |\widehat{\mathbf{c}}_j(\mu)|$ for the estimate $\widehat{\mathbf{c}}(\mu)$ produced in Step 2. An analogous threshold-based methodology was employed in [12].

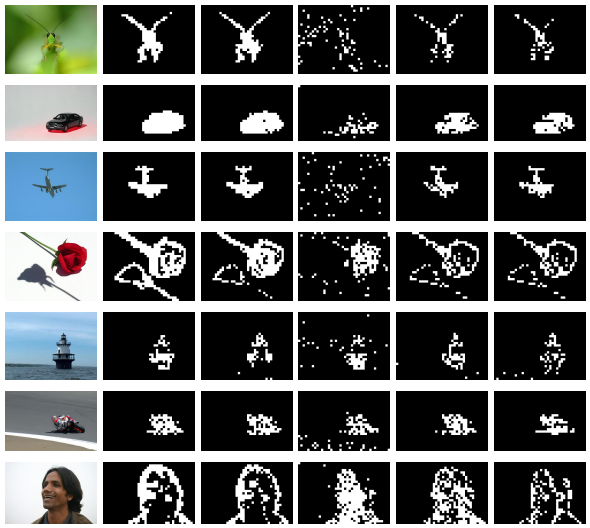


Fig. 3. Detection results for the MSRA Salient Object Database.

We see that increasing the parameter m of the matrix $\mathbf{\Phi}$ in Step 1 of our algorithm while keeping the other sampling parameters fixed (i.e., moving from top to bottom in any one column) facilitates accurate recovery for increasing ranks r of the matrix $\mathbf{L}$. Similarly, increasing the parameter p of the matrix $\mathbf{A}$ in Step 2 of our algorithm while keeping the other sampling parameters fixed (i.e., moving from left to right in any one row) facilitates accurate recovery for an increasing number k of outlier columns. Overall, our approach can successfully recover the locations of the outliers for non-trivial regimes of r and k using very few measurements.

We also compute phase transition curves for RMC using a similar methodology described above, with results provided in Figure 2. We observethat the RMC approach is viable for identifying the outliers from subsampled data provided the sampling rate exceeds about 10%. The relative difference in performance is likely due in large part to the difference in the observation models between the two approaches – the RMC approach is inherently operating in the presence of "missing data" while our approach permits us to observe linear combinations of any row or column of the entire matrix.

3.2. Real Data

We also evaluate the performance of our proposed method on real data in the context of a stylized image processing task that arises in many computer vision and automated surveillance applications – identifying the "saliency map" of an image. For this, we use images from the MSRA Salient Object Database [36].

Our approach here is based on representing each test image as a collection of non-overlapping image patches. We transform each (color) test image to gray scale, decompose it into non-overlapping 10×10-pixel patches, vectorize each patch into a 100×1 column vector, and assemble the column vectors into a matrix. For example, images of size 300×400 here yield matrices of size 100×1200. Notice that we only used gray scale values of image as the input feature rather than any high-level images feature – this facilitates the use of our approach, which is based on collecting linear measurements of the data (e.g., using a spatial light modulator, or an architecture like the *single pixel camera* [37]). We implement our ACOS method using two different sampling regimes, the first corresponding to $\gamma = 0.2$, $m = 0.2n_1$, $p = 0.5n_2$ (an average 4.5% sampling rate) and the other with $\gamma = 0.2$, $m = 0.1n_1$, $p = 0.5n_2$ (an average 2.5% sampling rate), and again generate $\mathbf{\Phi}$, $\mathbf{A}$, and ϕ to have iid Gaussian entries. We compare our approach with the OP approach (which uses the full data) and the RMC approach at sampling rates of 20% and 5%. The results are provided in Figure 3.

Table 1. Timing analysis for experiments on MRSA Database.

Method Sampling	OP 100%	RMC 20%	RMC 5%	ACOS 4.5%	ACOS 2.5%
Step 1	2.9441 (0.3854)	2.6324 (0.3237)	2.7254 (0.3660)	0.0533 (0.0118)	0.0214 (0.0056)
Step 2	– –	– –	– –	0.2010 (0.0674)	0.2014 (0.0692)

We note first that the OP approach, performs fairly well at identifying the visually salient regions, providing evidence to validate the use of the low-rank plus outlier model for visual saliency (see also [11]). Next, comparing the results of the individual procedures, we see that the OP approach appears to uniformly give the best detection results, which is reasonable since it is using the full data as input. The RMC approach performs well at the 20% sampling rate, but its performance appears to degrade at the 5% sampling rate. The ACOS approach, on the other hand, still produces reasonably accurate results using as low as a 2.5% sampling rate.

We also compare implementation times of the algorithms on this saliency map estimation task. Table 1 provides the average execution times (and standard deviations) for each approach, evaluated over 1000 images in the MSRA database[2]. Here, we only execute each procedure for one choice of regularization parameter. Overall, we see ACOS may be more than $15\times$ faster than the OP and RMC methods. At the expense of increased sample complexity, one could set $\mathbf{A} = \mathbf{I}$ in Step 2 of the ACOS method and thus eliminate the need to implement an iterative solver there. This could significantly speed up the second step of our method, and could result overall in relative speedups of up to $100\times$. Overall, our results suggest a significant improvement via ACOS for both detection consistency and timing, which may have a promising impact in salient signal detection tasks.

4. CONCLUSIONS

We proposed a novel two-step procedure to efficiently identify outlier columns embedded in a low-rank matrix using as few as $\mathcal{O}\left((r + \log k)(\mu_{\mathbf{L}} r \log r) + k \log(n_2/k)\right)$ linear measurements, and demonstrated its computational improvements over related approaches that operate on the "full data" directly. Further reductions in the complexity of our approach may be achieved using fast JL embeddings along the lines of [31, 32]. One interesting potential extension of our work would be to leverage our dimensionality reduction insight (focused only on outlier detection, not full recovery) within the context of the Compressive Principal Component Pursuit (Compressive PCP) of [38] in order to yield a procedure with comparable performance as ours, but which acquires only *non-adaptive* linear measurements of $\mathbf{M}$. Finally, we mention that our full-length version [33] of this work includes several extensions of the ideas presented here, including analysis of a "simplified" version of ACOS (where we choose $\mathbf{A} = \mathbf{I}$ in Step 2, to simplify the final inference), and evaluation of extensions to noisy and missing data settings.

[2]Timing comparisons were done with `MATLAB` R2013a on an iMac with a 3.4 GHz Intel Core i7 processor, 32 GB memory, and running OS X 10.8.5.

5. REFERENCES

[1] I. Jolliffe, *Principal component analysis*, Wiley Online Library, 2005.

[2] N. Halko, P.-G. Martinsson, and J. A. Tropp, "Finding structure with randomness: Probabilistic algorithms for constructing approximate matrix decompositions," *SIAM review*, vol. 53, no. 2, pp. 217–288, 2011.

[3] M. W. Mahoney, "Randomized algorithms for matrices and data," *Foundations and Trends® in Machine Learning*, vol. 3, no. 2, pp. 123–224, 2011.

[4] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] E. J. Candès and T. Tao, "Near optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[6] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," in *Matters of Intelligence*, pp. 115–141. Springer, 1987.

[7] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, 1998.

[8] J. Harel, C. Koch, and P Perona, "Graph-based visual saliency," in *Proc. NIPS*, 2006.

[9] N. Rao, J. Harrison, T. Karrels, R. Nowak, and T. T. Rogers, "Using machines to improve human saliency detection," in *Proc. Asilomar Conf. on Signals, Systems and Comp.*, 2010.

[10] G. Yu and G. Sapiro, "Statistical compressed sensing of Gaussian mixture models," *IEEE Trans. Signal Processing*, vol. 59, no. 12, pp. 5842–5858, 2011.

[11] X. Shen and Y. Wu, "A unified approach to salient object detection via low rank matrix recovery," in *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, 2012, pp. 853–860.

[12] H. Xu, C. Caramanis, and S. Sanghavi, "Robust PCA via outlier pursuit," *IEEE Trans. Inform. Theory*, vol. 58, no. 5, pp. 3047–3064, 2012.

[13] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9, pp. 589–592, 2008.

[14] E. Bashan, R. Raich, and A. O. Hero, "Optimal two-stage search for sparse targets using convex criteria," *IEEE Trans Signal Processing*, vol. 56, no. 11, pp. 5389–5402, 2008.

[15] J. Haupt, R. Castro, and R. Nowak, "Adaptive sensing for sparse signal recovery," in *Proc. IEEE DSP Workshop and Workshop on Sig. Proc. Education*, 2009, pp. 702–707.

[16] P. Indyk, E. Price, and D. P. Woodruff, "On the power of adaptivity in sparse recovery," in *Proc. IEEE Foundations of Computer Science*, 2011, pp. 285–294.

[17] M. Iwen and A. Tewfik, "Adaptive group testing strategies for target detection and localization in noisy environments," *IEEE Trans. Signal Proc.*, vol. 60, no. 5, pp. 2344–2353, 2012.

[18] M. Malloy and R. Nowak, "Near-optimal adaptive compressive sensing," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, 2012.

[19] M. A. Davenport and E. Arias-Castro, "Compressive binary search," in *Proc. IEEE Intl. Symp on Information Theory*, 2012, pp. 1827–1831.

[20] E. Arias-Castro, E. J. Candès, and M. A. Davenport, "On the fundamental limits of adaptive sensing," *IEEE Trans. Inform. Theory*, vol. 59, no. 1, pp. 472–481, 2013.

[21] A. Krishnamurthy and A. Singh, "Low-rank matrix and tensor completion via adaptive sampling," *arXiv preprint arXiv:1304.4672*, 2013.

[22] A. Soni and J. Haupt, "On the fundamental limits of recovering tree sparse vectors from noisy linear measurements," *IEEE Trans. Inform. Theory*, vol. 60, no. 1, pp. 133–149, 2014.

[23] J. Haupt and R. Nowak, "Adaptive sensing for sparse recovery," in *Compressed Sensing: Theory and applications*, Y. Eldar and G. Kutyniok, Eds. Cambridge University Press, 2011.

[24] V. Chandrasekaran, S. Sanghavi, P. Parrilo, and A. Willsky, "Rank-sparsity incoherence for matrix decomposition," *SIAM J. Optimization*, vol. 21, no. 2, 2011.

[25] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?," *J. of the ACM*, vol. 58, no. 3, 2011.

[26] Y. Chen, H. Xu, C. Caramanis, and S. Sanghavi, "Robust matrix completion with corrupted columns," *arXiv preprint arXiv:1102.2254*, 2011.

[27] M. McCoy and J. A. Tropp, "Two proposals for robust PCA using semidefinite programming," *Electronic Journal of Statistics*, vol. 5, pp. 1123–1160, 2011.

[28] G. Lerman, M. B. McCoy, J. A. Tropp, and T. Zhang, "Robust computation of linear models by convex relaxation," *Foundations of Computational Mathematics*, pp. 1–48, 2014.

[29] E. J. Candès and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational mathematics*, vol. 9, no. 6, pp. 717–772, 2009.

[30] A. C. Gilbert, J. Y. Park, and M. B. Wakin, "Sketched SVD: Recovering spectral features from compressive measurements," *arXiv preprint arXiv:1211.0361*, 2012.

[31] N. Ailon and B. Chazelle, "Approximate nearest neighbors and the fast Johnson-Llindenstrauss transform," in *Proc. ACM Symposium on Theory of Computing*, 2006, pp. 557–563.

[32] A. Dasgupta, R. Kumar, and T. Sarlós, "A sparse Johnson-Lindenstrauss transform," in *Proceedings ACM Symposium on Theory of Computing*, 2010, pp. 341–350.

[33] X. Li and J. Haupt, "Identifying outliers in large matrices via randomized adaptive compressive sampling," *arXiv preprint arXiv:1407.0312*, 2014, (accepted, to appear in IEEE Trans. Sig. Proc.).

[34] T. Goldstein, B. Donoghue, S. Setzer, and R. Baraniuk, "Fast alternating direction optimization methods," *CAM report 12-35, UCLA*, 2012.

[35] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.

[36] T. Liu, J. Sun, N. Zheng, X. Tang, and H. Shum, "Learning to detect a salient object," in *Proc. CVPR*, 2007.

[37] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *IEEE Sig. Proc. Magazine*, vol. 25, no. 2, pp. 83–91, 2008.

[38] J. Wright, A. Ganesh, L. Min, and Y. Ma, "Compressive principal component pursuit," *Information and Inference*, vol. 2, no. 1, pp. 32–68, 2013.