Compressive Independent Component Analysis

Michael P. Sheehan
Institute of Digital Communications
University of Edinburgh
Edinburgh, UK
michael.sheehan@ed.ac.uk

Madeleine S. Kotzagiannidis
Institute of Digital Communications
University of Edinburgh
Edinburgh, UK
madeleine.kotzagiannidis@ed.ac.uk

Mike E. Davies
Institute of Digital Communications
University of Edinburgh
Edinburgh, UK
mike.davies@ed.ac.uk

***Abstract*—In this paper we investigate the minimal dimension statistic necessary in order to solve the independent component analysis (ICA) problem. We create a compressive learning framework for ICA and show for the first time that the memory complexity scales only quadratically with respect to the number of independent sources n, resulting in a vast improvement over other ICA methods. This is made possible by demonstrating a low dimensional model set, that exists in the cumulant based ICA problem, can be stably embedded into a compressed space from a larger dimensional cumulant tensor space. We show that identifying independent source signals can be achieved with high probability when the compression size m is of the optimal order of the intrinsic dimension of the ICA parameters and propose a iterative projection gradient algorithm to achieve this.**

***Index Terms*—Compressive learning, random moments, compressive sensing, independent component analysis, statistical learning**

I. Introduction

Large scale learning is paramount to modern technology and research. A major drawback of many data analysis methods is that they need to store and return to the full data within their learning procedures, resulting in excessive storage demands and the need to store potentially sensitive data. Compressive learning alleviates these problems by compressing the data into a low dimensional sketch that captures enough information relevant for the learning task at hand [1]. The sketch is formed by a random set of summary statistics chosen a priori in a non adaptive fashion. The size of the sketch m, and therefore the dimension of the projected space, is chosen independently of the number of data samples T. This results in an algorithm which scales well with T, both computationally and in terms of the memory footprint. The size of the sketch m is set to be proportional to the dimension of the model set of the parametric model of interest. Therefore the scale of the compression is driven by the complexity of the task at hand. This strategy is akin to the techniques used in compressive sensing where noisy undetermined linear observations require regularization in the form of a low dimensional model set that the signal resides on or near to. With respect to compressive sensing, this model set usually takes the form of sparsity or low rank such that signal reconstruction is possible with limited observations. Cumulant based independent component analysis (ICA) exhibits such a low dimensional model set and therefore we can use this model set as a regularization to the ICA problem, enabling us to develop a compressive ICA framework.

This work was supported by the ERC Advanced grant, project C-SENSE, (ERC-ADG-2015-694888). Mike E. Davies is also supported by a Royal Society Wolfson Research Merit Award.

In this paper, we propose the scheme of compressive independent component analysis (CICA) and show that

1) A low dimensional model set exists within the cumulant based ICA problem.
2) Identifiability of the independent source signals is possible when the compressive sketch size is set proportional to the dimension of the model set.
3) The CICA framework is robust to a range of distributions and the memory demands are reduced substantially in comparison to existing techniques.

Firstly we define the cumulant based ICA model and the low dimensional model set.

A. Background

ICA is used frequently in the machine learning community to identify hidden independent factors within the data. Consider a data matrix $\mathbf{X} \in \mathbb{R}^{d\times T}$ where each row represents a feature and each column represents an individual sample. Let $\hat{\mathbf{x}} \in \mathbb{R}^d$ be an instance of $\mathbf{X}$, the goal of ICA is to identify the unknown mixing matrix $\mathbf{A} \in \mathbb{R}^{d\times n}$ up to inherent scaling and permutation ambiguities such that

$$\hat{\mathbf{x}} = \mathbf{A}\mathbf{s}, \tag{1}$$

under the assumption that the individual entries $s_1, s_2, \ldots, s_n$ of $\mathbf{s} \in \mathbb{R}^n$ are statistically independent:

$$p(s_1, s_2, \ldots, s_n) = \prod_{i=1}^{n} p_i(s_i). \tag{2}$$

For simplicity, we assume $d = n$ in this paper. The acquisition of these independent sources have been used to solve the blind source separation problem, as a dimensionality reduction technique and even to find hidden factors in financial data [2].

In general, most ICA methods prewhiten the data beforehand. This involves the process of finding the matrix $\mathbf{P}$ such that

$$\mathbf{x} = \mathbf{P}\hat{\mathbf{x}}, \tag{3}$$

where $\mathbf{x}$ has identity covariance matrix. This initial preprocessing step, which is executed through principal component analysis (PCA), makes the data uncorrelated. Moreover, it has

978-9-0827-9703-9/19/$31.00 ©2019 IEEE

the advantage that the matrix $\mathbf{Q} = \mathbf{PA}$ to be found is necessarily orthogonal. Throughout this paper we shall consider the whitened version of the data and the corresponding whitened ICA equation

$$\mathbf{x} = \mathbf{Q}\mathbf{s}. \tag{4}$$

There are many methods and algorithms in the literature to perform ICA. Tensorial or cumulant based methods are of particular interest here. Statistical properties of the data instance $\mathbf{x}$ can be described by its cumulants $\left(\mathcal{C}_{\mathbf{x}}^{N}\right)_{i_1 i_2 \dots i_N}$. Assuming the data has zero mean, the first four cumulants are defined [3] as

$$\begin{aligned}
\left(\mathcal{C}_{\mathbf{x}}^{1}\right)_{i} =& 0 \\
\left(\mathcal{C}_{\mathbf{x}}^{2}\right)_{ij} =& \delta_{ij} \\
\left(\mathcal{C}_{\mathbf{x}}^{3}\right)_{ijk} =& \mathbb{E}[x_i x_j x_k] \\
\left(\mathcal{C}_{\mathbf{x}}^{4}\right)_{ijkl} =& \mathbb{E}[x_i x_j x_k x_l] - \mathbb{E}[x_i x_j]\mathbb{E}[x_k x_l] \\
& - \mathbb{E}[x_i x_k]\mathbb{E}[x_j x_l] - \mathbb{E}[x_i x_l]\mathbb{E}[x_j x_k]
\end{aligned} \tag{5}$$

where $\mathbb{E}$ is the expectation operator. In the multivariate setting, cumulants give rise to tensors, denoted $\mathcal{C}_{\mathbf{x}}^{N}$ for a cumulant tensor of order N. The diagonal entries correspond to auto-cumulants and define the statistical properties of a distribution from a single component. On the other hand, the off-diagonals ($ijkl \neq iiii$) correspond to cross-cumulants and define the statistical dependencies between components from more than one distribution. The cross cumulants vanish if the components are statistically independent (defined in (2)), resulting in a strictly diagonal cumulant tensor. ICA works on the assumption that the n variables of $\mathbf{s}$ are statistically independent and that at most one source is Gaussian. As a result, the cumulant tensor formed by $\mathbf{s}$ is strictly diagonal with at most one zero on the leading diagonal.

Given the model in (4) equating $\mathbf{x}$ to $\mathbf{s}$, then the following multilinear property holds for their associated cumulant tensors [3]:

$$\mathcal{C}_{\mathbf{x}}^{N} = \mathcal{C}_{\mathbf{s}}^{N} \times_1 \mathbf{Q} \times_2 \mathbf{Q} \times_3 \cdots \times_N \mathbf{Q}, \tag{6}$$

where $\times_j$ represents the j-mode tensor-matrix product. In this paper we will only consider 4^{th} order cumulant tensors (e.g. $N = 4$) and for the sake of simplified notation we shall drop the superscript in equation (6) for the rest of the discussion. A comprehensive review of cumulants and tensors can be found in [3].

B. Model Set

Pioneered by Comon, Cardoso and De Lauthawer [4] [5] [6], the method of cumulant based ICA concerns diagonalising the cumulant tensor $\mathcal{C}_{\mathbf{x}}$ via an orthogonal transformation. In other words, these methods try to find a $\mathbf{Q}$ such that

$$\mathcal{C}_{\mathbf{s}} = \mathcal{C}_{\mathbf{x}} \times_1 \mathbf{Q}^T \times_2 \mathbf{Q}^T \times_3 \mathbf{Q}^T \times_4 \mathbf{Q}^T \tag{7}$$

is strictly diagonal. The set of cumulant tensors $\mathfrak{C}$ which can be orthogonally transformed into a diagonal tensor creates a low dimensional model set residing in the whole tensor space. Formally, this model set can be defined as

$$\begin{aligned}
\mathfrak{S} := \{\mathcal{C}_{\mathbf{x}} \in \mathfrak{C} : \mathcal{C}_{\mathbf{x}} \times_1 \mathbf{Q}^T \times_2 \mathbf{Q}^T \times_3 \\
\mathbf{Q}^T \times_4 \mathbf{Q}^T \in \mathcal{T}_{\text{diag}}, \mathbf{Q}^T\mathbf{Q} = \mathbf{I}_n\}
\end{aligned} \tag{8}$$

where $\mathcal{T}_{\text{diag}}$ is the set of strictly diagonal tensors such that $(\mathcal{C})_{ijkl} = 0$ for all $ijkl \neq iiii$. The model set $\mathfrak{S}$ has dimension $\frac{n(n+1)}{2}$ which can be easily deduced:

- n variables (or degrees of freedom) on the diagonal of the cumulant tensor $\mathcal{C}_{\mathbf{s}}$;
- $\frac{n(n-1)}{2}$ variables on the orthogonal matrix $\mathbf{Q}$;
- total degrees of freedom: $\frac{n(n-1)}{2} + n = \frac{n(n+1)}{2}$

The intrinsic dimension of the model set $\mathfrak{S}$ is $|\mathfrak{S}| = \frac{n(n+1)}{2} \approx \mathcal{O}(n^2)$ in comparison to the dimension of the set of 4^{th} order cumulant tensors, scaling at $|\mathfrak{C}| = \mathcal{O}(n^4)$. Consequently, the model set has intrinsic dimension far smaller than the space it resides in. The realisation of a model set within the cumulant based ICA enables us to develop a compressive learning framework for ICA.

II. Related Work

Compressive learning was pioneered by Keriven and Gribonval [1] [7] [8] and builds on the existing fields of linear sketching and randomized kernel approximation. Related to the work developed in this paper, a compressive principal component analysis (PCA) framework is developed in [1] that exploits the low rankness of the covariance matrix of the data to create a model set to sketch upon. In PCA, the covariance matrix of the data admits a finite dimensional sufficient statistic as it summarizes all the information needed to select a best subspace of maximal variance. ICA is an interesting and more challenging model as it does not admit a minimal sufficient statistic and for that reason the use of cumulant tensors is leveraged to act as an intermediary polynomial map to a finite, yet large, dimensional space. The ICA problem is generally identifiable through the cumulant statistics. However, their size is not order optimal. Here we show that this can be rectified via our compressive learning framework.

In other related work, Sela et al [9] used kernel approximation techniques to reduce the dimensions of the Kernel ICA method proposed by Bach [10]. More specifically random Fourier features are used to approximate the kernel, reducing the memory complexity from $\mathcal{O}(n^2T^2)$ to $\mathcal{O}(\omega T)$, where ω is the number of random Fourier weights used. Despite the reduction in memory complexity, the algorithm still has storage demands which scale linearly with T. In comparison, we remove the dependency of the data length T completely, within our framework, when estimating the mixing matrix.

III. Compressive ICA

In section I we established a model set $\mathfrak{S}$ which is a subset of the space of cumulant tensors $\mathfrak{C}$ where the solution to the

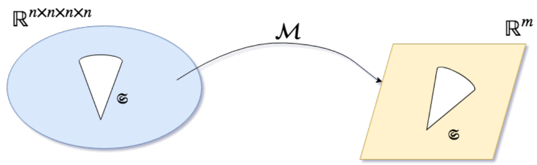


Fig. 1: A schematic diagram of the sketching operator $\mathcal{M}$.

ICA problem resides. The model set $\mathfrak{S}$ has intrinsic dimension equal to $\frac{n(n+1)}{2}$ which is far smaller than the $\mathcal{O}(n^4)$ dimension of the space of cumulant tensors. In theory, this enables us to embed the model set $\mathfrak{S}$ from the space of 4^{th} order cumulant tensors $\mathfrak{C}$ it resides in, to a lower dimensional embedded space with minimal distortion. Compressive learning theory [1] [11] states that robust linear embeddings are achievable through random projections when the embedded space is of the order of the model set $\mathfrak{S}$. In other words, we can construct a random linear mapping $\mathcal{M} : \mathfrak{C} \to \mathbb{R}^m$ such that the embedded space has dimension $m = \mathcal{O}(|\mathfrak{S}|) \approx \mathcal{O}(n^2)$ and distances are approximately preserved on the model set. We therefore define the sketching operator as

$$\begin{aligned} \mathcal{M} :& \mathfrak{C} \to \mathbb{R}^m \\ & \mathcal{C}_{\mathbf{x}} \mapsto \mathbf{M}\text{vec}(\mathcal{C}_{\mathbf{x}}) \end{aligned} \tag{9}$$

where $\mathbf{M} \in \mathbb{R}^{m\times n^4}$ is a subgaussian matrix with entries $\mathbf{M}_{ij} \sim \mathcal{N}(0, 1/m)$ and vec is the vector operator that restructures the cumulant tensor as a vector. A schematic diagram of the compressive sketching operator $\mathcal{M}$ is illustrated in figure 1.

Equipped with the sketching operator , we define the compressive sketch as

$$\mathbf{y} = \mathcal{M}(\mathcal{C}_{\mathbf{x}}). \tag{10}$$

Once the compressive sketch $\mathbf{y}$ is constructed, we have to infer the parameters of the ICA model solely from the sketch. This results in the following linear inverse optimisation problem to be solved

$$\min_{\mathcal{C}_{\mathbf{x}} \in \mathfrak{S}} \|\mathbf{y} - \mathcal{M}(\mathcal{C}_{\mathbf{x}})\|_2 \tag{11}$$

Equation (11) represents an undetermined linear inverse problem with the constraint that the solution belongs to the model set $\mathfrak{S}$. This constraint acts as a regularizer ensuring the problem is not ill-posed and therefore a unique solution can be found. In this light, we propose using an iterative projection gradient (IPG) algorithm [12] [13] to find the minimiser of (11). The IPG employs the following iterative scheme

$$\mathcal{C}_{\mathbf{x}}^{j+1} = \mathcal{P}_{\mathfrak{S}}\Big(\mathcal{C}_{\mathbf{x}}^{j} + \lambda\mathcal{M}^T\big(\mathbf{y} - \mathcal{M}(\mathcal{C}_{\mathbf{x}}^{j})\big)\Big), \tag{12}$$

where $\mathcal{P}_{\mathfrak{S}}$ is a projection operator, projecting the updated cumulant tensor onto the model set $\mathfrak{S}$ at each iteration, λ is the step size, and $\mathcal{M}^T(\cdot) = \text{unvec}(\mathbf{M}^T\cdot)$ is the adjoint sketching operator where unvec is the adjoint vector operator.

A. Projection Operator $\mathcal{P}_{\mathfrak{S}}$

In general, finding a computationally tractable projection operator onto a model set is complex and non trivial. As this is the case with compressive ICA, we employ a Jacobi diagonalisation followed by hard thresholding to act as a proxy projection operator. The Jacobi diagonalisation consists of maximising a contrast function via consecutive Givens rotations. Constrast functions are used in ICA methods and give a quantifiable and computationally tractable measure of statistical independence of a given cumulant tensor. In the compressive ICA case, we maximimise the following contrast function with respect to the matrix $\mathbf{Q}$:

$$\Psi(\mathbf{Q}) = \Big(\mathcal{C}_{\mathbf{x}} \times_1 \mathbf{Q}^T \times_2 \mathbf{Q}^T \times_3 \mathbf{Q}^T \times_4 \mathbf{Q}^T\Big)^2_{iiii}. \tag{13}$$

The resulting Jacobi scheme produces an approximately diagonalised 4^{th} order cumulant tensor and therefore we apply a hard thresholding operator by forcing the off-diagonals of $\mathcal{C}_{\mathbf{x}} \times_1 \mathbf{Q}^T \times_2 \mathbf{Q}^T \times_3 \mathbf{Q}^T \times_4 \mathbf{Q}^T$ to zero. This procedure projects the updated cumulant tensor in (12) on to the model set $\mathfrak{S}$ at each iteration and therefore acts as a proxy projector. In section IV, we see this scheme works very well in practice, in terms of stability and performance, and that it is also computationally tractable.

The IPG algorithm defined in (12) has the following parameters:

- Initial 4^{th} order cumulant tensor: $\mathcal{C}_{\mathbf{x}}^0 = \mathbf{0}$
- Stopping criteria: when $\|\mathbf{y} - \mathcal{M}(\mathcal{C}_{\mathbf{x}})\|_2 \leq 10^{-5}$
- Step size : $\lambda_j = \frac{\|\mathcal{M}^T(\mathbf{y} - \mathcal{M}(\mathcal{C}_{\mathbf{x}}^j))\|_F^2}{\|\mathcal{M}(\mathcal{M}^T(\mathbf{y} - \mathcal{M}(\mathcal{C}_{\mathbf{x}}^j)))\|_2^2}$

The compressive ICA algorithm converges very fast at around 10-20 iterations in practice.

The IPG scheme in (12) uses only the compressive statistic $\mathbf{y}$ throughout and never needs to return to the original data $\mathbf{X}$ nor the cumulant tensor $\mathcal{C}_{\mathbf{x}}$ at any stage. This means we only need to store the compressive sketch $\mathbf{y}$ in memory, which does not scale with the size of the data T, resulting in substantial memory footprint savings. Table I highlights the memory complexity for several batch ICA techniques found in the literature: Joint Approximation Diagonalization of Eigenmatrices (JADE) [5], Fast ICA (FICA) [2], Kernel ICA (KICA) [10] and Comon's ICA algorithm (ComICA) [4].

ICA Method	Memory Complexity
FICA	$\mathcal{O}(nT + n^2)$
ComICA	$\mathcal{O}(nT + n^2)$
JADE	$\mathcal{O}(n^4)$
KICA [1]	$\mathcal{O}(rT + n^2)$
Compressive ICA	$\mathcal{O}(n^2)$

TABLE I: A table showing the memory complexities of batch ICA algorithms for estimating the mixture matrix $\mathbf{A}$.

In the next section we will run extensive experiments on a range of distributions to show that such stable compression is

[1] Incomplete Cholesky factorization is used in [10] to reduce memory complexity from $\mathcal{O}(T^2)$ to $\mathcal{O}(rT)$ where $r \ll T$ is the rank approximation of the Gram matrix.

possible and to numerically show that the order of magnitude in the compression $m = \mathcal{O}(n^2)$ is approximately in the region 2-3.

IV. Numerical Experiments

Throughout this section the Amari error [14], defined below, will be used to evaluate the accuracy of the mixing matrix estimate. Given the mixing matrix $\mathbf{A}$ and the mixing matrix estimate $\hat{\mathbf{A}}$, the Amari error is defined by

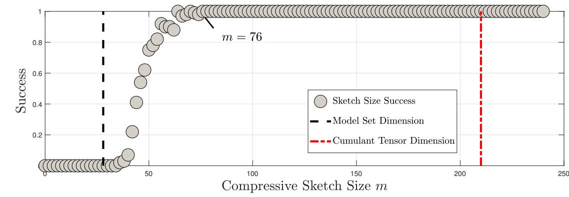
$$\begin{aligned} d(\mathbf{A}, \hat{\mathbf{A}}) =& \frac{1}{2n} \sum_{i=1}^{n} \left(\frac{\sum_{j=1}^{n} |b_{ij}|}{\max_j |b_{ij}|} - 1 \right) \\ &+ \frac{1}{2n} \sum_{j=1}^{n} \left(\frac{\sum_{i=1}^{n} |b_{ij}|}{\max_i |b_{ij}|} - 1 \right), \end{aligned} \tag{14}$$

where $b_{ij} = (\mathbf{A}\hat{\mathbf{A}}^{-1})_{ij}$. The Amari error is used widely as it is both scale and permutation invariant and so therefore can be used to compare ICA mixing matrix estimates. The Amari error is bounded between 0 and 1 where $d(\mathbf{A}, \hat{\mathbf{A}}) = 0$ if and only if $\mathbf{A} = \hat{\mathbf{A}}$ up to scaling and permutation ambiguities.

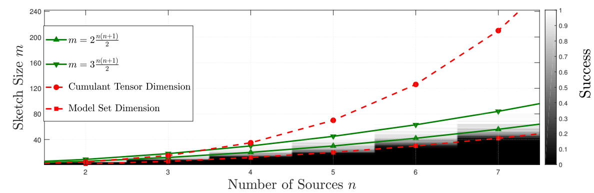
In the following subsection we construct experiments to determine the size of the sketch needed in order to solve the ICA problem. More specifically, we determine the order of magnitude in the sketch size $m = \mathcal{O}(\frac{n(n+1)}{2})$. Secondly, we show that the average error reduces optimally as the signal length T increases for different sketch sizes. In other words, the statistical efficiency of our compressive ICA estimates are optimal.

Firstly, to determine the size of the sketch needed for successful signal reconstruction, we construct a phase transition test. The phase transition is common in the compressive sensing literature [15] and highlights a sharp transition to a state of successful reconstruction as the compression size m increases. The phase transition empirically shows a lower bound region to the compression size m. To set up the phase transition experiment, the analytic cumulant tensor of Laplacian distributions were transformed by a known orthogonal mixing matrix $\mathbf{Q}$ as in (6). For each n, we ran tests on the compressive sketch size ranging from $m = 0$ to $m = 240$. For each m, 100 separate tests were replicated and success was determined by inspecting if the Amari error between the known mixing matrix $\mathbf{Q}$ and the estimate $\hat{\mathbf{Q}}$ was below the tolerance $d(\mathbf{Q}, \hat{\mathbf{Q}}) \leq 10^{-6}$. The results are illustrated in Figure 2.

The results clearly show a sharp transition in successful reconstruction for all n. The superimposed lines suggest that the compressive sketch size should be chosen to be greater than 2 times the model set dimension $\mathfrak{S}$ so that an accurate estimate of the mixing matrix is recovered with high probability. By choosing an order of magnitude of $c = 2$, we get a compression of 73% when $n = 7$. Interestingly, the compression in fact increases as n grows larger, indicating the upper full data bound increases at a faster rate than the sketch size bound. In figure 2, the lower red line highlights the dimensions of the model set $\mathfrak{S}$, illustrating the absolute lower bound to the ICA problem. It is therefore informative to see that the compression size m needed for successful reconstruction scales proportionally to the lower bound. These results indicate that the compression size m is of optimal order of the intrinsic dimension of the ICA parameters and model set $\mathfrak{S}$.



(a)



(b)

Fig. 2: (a) Single phase transition for $n = 7$ unit scale Laplacian sources. (b) Phase transition for $n = 2$ to $n = 7$ Laplacian sources.

In this next experiment we consider $n = 5$ independent sources sampled from Laplacian distributions and mixed with an known mixing matrix $\mathbf{A}$. The accuracy of the mixing matrix estimate $\hat{\mathbf{A}}$ was assessed using the Amari error and averaged over 1000 runs for each compression size m and signal length T. Figure 3 plots the $\log_{10}$ signal length T against the $\log_{10}$ average Amari error.

It is well documented [2] in estimation theory that statistical estimators follow the central limit theorem asymptotically as the data size T increases. Formally, the ICA error scales

$$d(\mathbf{A}, \hat{\mathbf{A}}) = \mathcal{O}\Big(1/\sqrt{T}\Big). \tag{15}$$

Figure 3 shows for each optimal compressive size m (i.e $m = c\frac{n(n+1)}{2}$ for $c > 2$ seen in the previous experiment) the $\log_{10}$ mean Amari error reduces at a rate of approximately $-1/2$ with respect to $\log_{10}(T)$. This clearly shows that the optimal

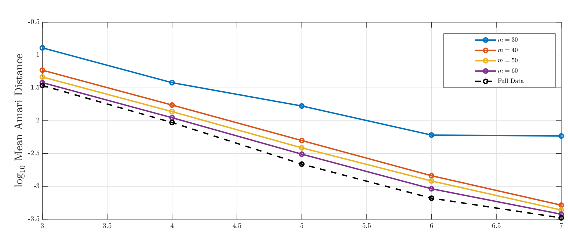


Fig. 3: A graph showing the $\log_{10}$ mean Amari error plotted against the $\log_{10}$ signal length T to show the efficiency of the compressive ICA estimates and the full data estimate.

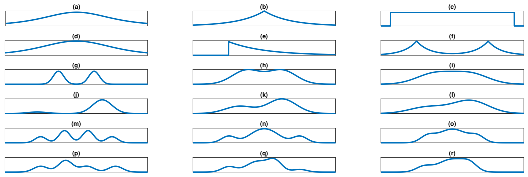


Fig. 4: Probability density functions of sources: (a) Student with 3 degrees of freedom; (b) double exponential; (c) uniform; (d) Student with 5 degrees of freedom; (e) exponential; (f) mixture of two double exponentials; (g)-(h)-(i) symmetric mixtures of two Gaussians: multimodal, transitional and unimodal; (j)-(k)-(l) nonsymmetric mixtures of two Gaussians, multimodal, transitional and unimodal; (m)-(n)-(o) symmetric mixtures of four Gaussians: multimodal, transitional and unimodal; (p)-(q)-(r) nonsymmetric mixtures of four Gaussians: multimodal, transitional and unimodal. The code used to create the figures can be found in [10]

Number of Sources n	Amari Error	Memory Compression
4	0.0147	15%
5	0.0502	36%
6	0.0359	50%
7	0.0691	60%
8	0.0482	67%
9	0.0826	73%
10	0.0563	78%

TABLE II: A table showing the average Amari error of 100 replicates where the source distributions are uniformly randomly chosen over the 18 distributions illustrated in figure 4. The last column highlights the amount of memory compression in relation to a full data cumulant based approach.

$\mathcal{O}\left(1/\sqrt{T}\right)$ learning rate exhibited by the full data case is not sacrificed when using compressive ICA. Moreover, the compressive size $m = 30$ is not an optimal compression size and as expected has not produced optimal order learning rates.

A. Experiments 2

In this section we conduct extensive experiments on sub, super and nearly Gaussian distributions to show the robustness of our framework and to highlight the significant compression compared to existing techniques in the ICA literature. In this experiment, we mix n sources randomly chosen from the 18 sources illustrated in figure 4 of length $T = 10000$ by a random mixing matrix $\mathbf{A}$ and present the Amari error and the associated memory compression compared to the full data cumulant based approach (i.e no compression).

The results indicate that the compressive ICA method is robust to a range of distributions. The experiment consisted of selecting n sources distributed uniformly at random from the 18 pdfs illustrated in figure 4. The average Amari error is consistently low and close to zero throughout each n-way ICA experiment. Moreover, we have highlighted the memory compression for each n-way ICA problem in comparison to the no compression case i.e. m is equal to the dimension of the space of 4^{th} order cumulant tensors $\mathfrak{C}$. The memory compression increases as n increases therefore making compressive ICA amenable to large scale n-way ICA problems.

V. Conclusion

In this paper we have shown that the cumulant based ICA method has a model set $\mathfrak{S}$ with intrinsic dimensions far smaller than the cumulant tensor space it resides in. By employing randomised techniques, we show that $\mathfrak{S}$ can be stably embedded into a space of dimension m which is empirically shown to be 2-3 times the intrinsic dimension of $\mathfrak{S}$. This results in substantial compression which increases as the number of sources n increases. We show that the CICA method leads to impressive compression in memory footprint, resulting in a fraction of the memory needed when compared to other ICA methods. In future work we will investigate if it is possible to work directly in the compressed measurement domain $\mathbb{R}^m$ to reduce the computation complexity of the ICA problem. Also we will derive RIP bounds on the sketching operator $\mathcal{M}$ and derive the theoretical efficiency associated with the compressive sketch size m which we saw in figure 3.

References

[1] R. Gribonval, G. Blanchard, N. Keriven, and Y. Traonmilin, "Compressive statistical learning with random feature moments," *arXiv preprint arXiv:1706.07180*, 2017.
[2] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural networks*, vol. 13, no. 4-5, pp. 411–430, 2000.
[3] L. De Lathauwer, *Signal processing based on multilinear algebra*, Katholieke Universiteit Leuven Leuven, 1997.
[4] P. Comon, "Independent component analysis, a new concept?," *Signal processing*, vol. 36, no. 3, pp. 287–314, 1994.
[5] J.F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," in *IEE proceedings F (radar and signal processing)*. IET, 1993, vol. 140, pp. 362–370.
[6] L. De Lathauwer, B. De Moor, and J. Vandewalle, "An introduction to independent component analysis," *Journal of Chemometrics: A Journal of the Chemometrics Society*, vol. 14, no. 3, pp. 123–149, 2000.
[7] N. Keriven, A. Bourrier, R. Gribonval, and P. Pérez, "Sketching for large-scale learning of mixture models," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 6190–6194.
[8] Nicolas Keriven, "Sketching for Large-Scale Learning of Mixture Models To cite this version : HAL Id : tel-01620815 Nicolas Keriven Sketching for Large- Scale Learning of Mixture Models," 2017.
[9] M. Sela and R. Kimmel, "Randomized independent component analysis," in *Science of Electrical Engineering (ICSEE), IEEE International Conference on the Science of Electrical Engineering*. IEEE, 2016, pp. 1–5.
[10] F.R. Bach and M.I. Jordan, "Kernel independent component analysis," *Journal of machine learning research*, vol. 3, no. Jul, pp. 1–48, 2002.
[11] G. Puy, M.E. Davies, and R. Gribonval, "Recipes for stable linear embeddings from Hilbert spaces to $\mathbb{R}^m$," *IEEE Transactions on Information Theory*, vol. 63, no. 4, pp. 2171–2187, 2017.
[12] T. Blumensath and M.E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and computational harmonic analysis*, vol. 27, no. 3, pp. 265–274, 2009.
[13] Thomas Blumensath, "Sampling and reconstructing signals from a union of linear subspaces," *CoRR*, vol. abs/0911.3514, 2009.
[14] S. Amari, A. Cichocki, and H. Yang, "A new learning algorithm for blind signal separation," in *Advances in neural information processing systems*, 1996, pp. 757–763.
[15] Y.C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*, Cambridge University Press, 2012.