Compressed Sensing and Source Separation

Thomas Blumensath and Mike Davies ⋆

IDCOM & Joint Research Institute for Signal and Image Processing
The University of Edinburgh, The King's Buildings, Edinburgh, EH9 3JL, UK
Tel.: +44(0)131 6505659, Fax.: +44(0)131 6506554, e-mail:
thomas.blumensath@ed.ac.uk, mike.davies@ed.ac.uk.

Abstract. Separation of underdetermined mixtures is an important problem in signal processing that has attracted a great deal of attention over the years. Prior knowledge is required to solve such problems and one of the most common forms of structure exploited is sparsity.
Another central problem in signal processing is sampling. Recently, it has been shown that it is possible to sample well below the Nyquist limit whenever the signal has additional structure. This theory is known as compressed sensing or compressive sampling and a wealth of theoretical insight has been gained for signals that permit a sparse representation.
In this paper we point out several similarities between compressed sensing and source separation. We here mainly assume that the mixing system is known, i.e. we do not study *blind* source separation. With a particular view towards source separation, we extend some of the results in compressed sensing to more general overcomplete sparse representations and study the sensitivity of the solution to errors in the mixing system.

1 Compressed Sensing

Compressed sensing or compressive sampling is a new emerging technique in signal processing, coding and information theory. For a good place of departure see for example [1] and [2]. Assume that a signal $\mathbf{y}$ is to be measured. In general $\mathbf{y}$ is assumed to be a function defined on a continuous domain, however, for the discussion here it can be assumed to be a finite vector, i.e. $\mathbf{y} \in \mathbb{R}^{N_y}$ say. In a standard DSP textbook we learn that one has to sample a function on a continuous domain at least at its Nyquist rate. However, assume that we know that $\mathbf{y}$ has a certain structure, for example we assume that $\mathbf{y}$ can be expressed as

$$\mathbf{y} = \mathbf{\Phi s}, \tag{1}$$

where $\mathbf{\Phi} \in \mathbb{R}^{\mathbf{N}_y \times \mathbf{N}_s}$ and where we allow $\mathbf{N}_s \geq \mathbf{N}_y$, i.e. we allow $\mathbf{\Phi s}$ to be an *overcomplete* representation of $\mathbf{y}$. Crucially, we assume $\mathbf{s}$ to be sparse, i.e. we assume that only a small number of elements in $\mathbf{s}$ are non-zero or, more generally,

⋆ This research was supported by EPSRC grant D000246/1. MED acknowledges support of his position from the Scottish Funding Council and their support of the Joint Research Institute with the Heriot-Watt University as a component part of the Edinburgh Research Partnership.

that most of the energy in $\mathbf{s}$ is concentrated in a few coefficients. It has recently been shown that, if a signal has such a sparse representation, then it is possible to take less samples (or measurements) from the signal than would be suggested by the Nyquist limit. Furthermore, one is then often still able to reconstruct the original signal using convex optimisation techniques [1] and [2].

The simplest scenario are measurements taken as follows:

$$\mathbf{x} = \mathbf{M}\mathbf{y}, \tag{2}$$

where $\mathbf{x} \in \mathbb{R}^{N_x}$ with $N_x < N_y$. Extensions to noisy measurements can be made [2], i.e. one can consider the problem of approximating $\mathbf{y}$ given a noisy measurement:

$$\tilde{\mathbf{x}} = \mathbf{x} + \mathbf{e} = \mathbf{M}\mathbf{y} + \mathbf{e}. \tag{3}$$

The ability to reconstruct the original signal relies heavily on the structure of $\mathbf{y}$ and different conditions have been derived under which one can exactly or approximately recover $\mathbf{y}$. For example, if $\mathbf{y} = \mathbf{\Phi}\mathbf{s}$ and $\mathbf{s}$ has only a small number of non-zero elements, then linear programming can exactly recover $\mathbf{y}$ if enough measurements have been taken. Similar results have been derived in the case where $\mathbf{s}$ is not exactly sparse, but where the ordered coefficients in $\mathbf{s}$ decay with a power law. In this case $\mathbf{y}$ can be recovered up to some small error. We give examples of these theorems below. More details can be found in for example [2] and the references therein.

2 Relationship to Source Separations

Sparsity has also often been exploited for source separation. In particular, the problem of underdetermined blind source separation has been solved using the fact that an orthogonal transform can often be found in which the data is sparse [3] [4] [5] [6]. More general, possibly over-complete dictionaries have been used for source separation in [7].

Let us assume a quite general source separation scenario. A set of sources, say $\mathbf{g}_1, \mathbf{g}_2, \ldots$, each represented in a column vector, are collected into a matrix

$$\mathbf{G} = [\mathbf{g}_1 \ \mathbf{g}_2 \ \ldots]^T \tag{4}$$

and similarly, the set of observations $\mathbf{f}_1, \mathbf{f}_2, \ldots$ are gathered in a matrix $\mathbf{F}$. The relationship between the sources $\mathbf{G}$ and the observations F is then modelled by a general linear operator $\mathbf{A}$:

$$\mathbf{F} = \mathbf{A}(\mathbf{G}). \tag{5}$$

Note that this operator does not have to be a matrix and can also represent for example convolutions, so that the above model incorporates a wide range of source separation problems.

The connection to compressive sampling becomes evident if instead of collecting the sources and observations in a matrix, we interleave them into vectors as:

$$\mathbf{x} = [\mathbf{f}_1[1] \ \mathbf{f}_2[1] \ \ldots \ \mathbf{f}_1[2] \ \mathbf{f}_2[2] \ \ldots]^T \tag{6}$$

and

$$\mathbf{y} = [\mathbf{g}_1[1]\ \mathbf{g}_2[1]\ \dots\ \mathbf{g}_1[2]\ \mathbf{g}_2[2]\ \dots]^T . \tag{7}$$

We further assume that the operator $\mathbf{A}$ can be expressed in matrix form $\mathbf{M}$ so that the mixing system becomes:

$$\mathbf{x} = \mathbf{M}\mathbf{y}, \tag{8}$$

which is exactly the compressive sampling measurement equation[1].

If we have more sources $\mathbf{g}$ than observations $\mathbf{f}$, where the length of each observation and each source is assumed to be equal, then we have less measurements than samples in $\mathbf{y}$. We therefore require knowledge of additional structure if we want to be able to (approximately) reconstruct $\mathbf{y}$. We can, for example, assume the existence of a sparse representation of $\mathbf{y}$ of the form $\mathbf{y} = \mathbf{\Phi}\mathbf{s}$, where $\mathbf{s}$ is sparse. Note that we do not assume that $\mathbf{\Phi}$ is an orthogonal transform and explicitly allow $\mathbf{\Phi}$ to be overcomplete, i.e. to have more columns than rows.

Let us look at a simple example of an instantaneous mixture. In this case, $\mathbf{A}$ is the $N_f \times N_g$ mixing matrix and the matrix $\mathbf{M}$ becomes matrix diagonal:

$$\mathbf{M} = \begin{bmatrix} \mathbf{A} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{A} & \cdots & \mathbf{0} \\ \vdots & & \ddots & \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A} \end{bmatrix}, \tag{9}$$

Where $\mathbf{0}$ is a $N_f \times N_g$ matrix of zeros. Similarly, assume a convolutive model, in which the impulse responses, say $\mathbf{h}_{1,1}, \mathbf{h}_{2,1}, \mathbf{h}_{3,1}$ and $\mathbf{h}_{1,2}, \mathbf{h}_{2,2}, \mathbf{h}_{3,2}$ and so on are interleaved into the matrix $\mathbf{H}$ as follows:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_{1,1}[n]\ \mathbf{h}_{2,1}[n]\ \mathbf{h}_{3,1}[n]\ \mathbf{h}_{1,1}[n-1] \cdots \\ \mathbf{h}_{1,2}[n]\ \mathbf{h}_{2,2}[n]\ \mathbf{h}_{3,2}[n]\ \mathbf{h}_{1,2}[n-1] \cdots \\ \vdots \end{bmatrix}. \tag{10}$$

The measuring matrix then becomes:

$$\mathbf{M} = \begin{bmatrix} \mathbf{H} & & \cdots \mathbf{0} \\ \mathbf{0} & \mathbf{H} & \cdots \mathbf{0} \\ \vdots & & \ddots \\ \mathbf{0} & \mathbf{0} & \cdots \mathbf{H} \end{bmatrix}, \tag{11}$$

where again $\mathbf{0}$ is a $N_f \times N_g$ matrix of zeros. Also, depending on boundary assumptions, the first and last rows of $\mathbf{M}$ might only contain part of the matrix $\mathbf{H}$.

[1] We could have alternatively stacked the vectors $\mathbf{f}_1, \mathbf{f}_2, \dots$ (and/or $\mathbf{g}_1, \mathbf{g}_2, \dots$) on top of each other to produce a permutation of the above model.

3 Theoretic Results

The source separation problem is equivalent to the decoding problem faced in compressive sampling and theoretical results from the compressive sampling literature therefore also apply to the source separation problem. However, in compressive sampling, the measurement matrix $\mathbf{M}$ can often be 'designed' to fulfil certain conditions[2]. Furthermore, in the current compressive sampling literature, $\mathbf{\Phi}$ is normally assumed to be the identity matrix or an orthogonal transform. In source separation, the measuring system is not normally at our control. Furthermore, orthogonal transform are often not available to sufficiently 'sparsify' many signals of interest.

In this paper we address these problems and extend several results from the compressed sensing literature to more general sparse representations. We start by reviewing some of the important results on compressed sensing [8], which we here write in terms of the m-restricted isometry condition of the matrix $\mathbf{P} = \mathbf{M\Phi}$.

3.1 *m*-restricted isometry

For any matrix $\mathbf{P}$ and integer m, define the m-restricted isometry $\delta_m(\mathbf{P})$ as the smallest quantity such that:

$$(1 - \delta_m(\mathbf{P})) \leq \frac{\|\mathbf{P}_\Gamma \mathbf{y}\|_2^2}{\|\mathbf{y}\|_2^2} \leq (1 + \delta_m(\mathbf{P})), \tag{12}$$

for all $\Gamma : |\Gamma| \leq m$ and all $\mathbf{y}$. Here $|\Gamma|$ is a set of indices and $\mathbf{P}_\Gamma$ the associated submatrix of $\mathbf{P}$ with all columns removed apart from those with indices in Γ. δ_m is then a measure of how much any sub-matrices of $\mathbf{P}$ with size m can change the norm of a vector, hence the name. The quantities $(1 - \delta_m(\mathbf{P}))$ and $(1 + \delta_m(\mathbf{P}))$ can be understood as lower and upper bounds on the squared singular values of all possible sub-matrices of $\mathbf{P}$ with m or less columns.

3.2 Estimation error bounds

As examples of the types of theorems available in the compressive sampling literature, we here state two of the fundamental results (these can be found in [2] and references therein), which rely on $\delta_m(\mathbf{M\Phi})$ to be small[3].

[2] 'Design' here often means taking a random matrix drawn from certain distributions.

[3] Note that Georgiev et al. [9] have also studied a similar problem in relationship to blind source separation. However, the results in [9] are concerned with identifiability of both $\mathbf{y}$ and $\mathbf{A}$. The theorems given here assume knowledge of $\mathbf{A}$ but are stronger than those in [9] in that they also states that we can use convex optimisation methods to identify the sources. Furthermore, theorem 2 is valid for more general sources and does not require the existence of an exact m-term representation.

Theorem 1 *(Exact Recovery) Assume that* $\mathbf{s}$ *has a maximum of* m *non-zero coefficients and that* $\mathbf{x} = \mathbf{M\Phi s}$ *and that* $\delta_{2m}(\mathbf{M\Phi}) + \delta_{3m}(\mathbf{M\Phi}) < 1$*, then the solution to the linear program :*

$$\min \|\hat{\mathbf{s}}\|, \text{ such that } \mathbf{x} = \mathbf{M\Phi}\hat{\mathbf{s}} \tag{13}$$

recovers the exact representation $\mathbf{y} = \mathbf{\Phi s}$.

A similar result can be derived for noisy observations and a further generalisation was derived in [8] for signals for which the original signal is not m-sparse, but has a power law decay, i.e. the magnitude of the ordered coefficients decays as $|s_{i_k}| \leq Ck^{-\frac{1}{p}}$, where $p \leq 1$. In particular, for an i.i.d. Gaussian observation error with variance σ^2, we have:

Theorem 2 *(Dantzig selector) Assume that* $\mathbf{s}$ *can be reordered so that* $|s_{i_k}| \leq ck^{-\frac{1}{p}}$ *for* $p \leq 1$*. For some* m *assume that* $\delta_{2m}(\mathbf{M\Phi}) + 3\delta_{3m}(\mathbf{M\Phi}) < 1$*. For* $\lambda = \sqrt{2\log N_x}$ *the solution to:*

$$\min \|\hat{\mathbf{y}}\|_1 \; : \; \|\mathbf{M}^T(\mathbf{x} - \mathbf{M}\hat{\mathbf{y}})\|_\infty \leq \lambda\sigma \tag{14}$$

obeys the bound:

$$\|\mathbf{s} - \tilde{\mathbf{s}}\|_2^2 \leq C2(\log N_x)\sigma^p c^{1-\frac{p}{2}}. \tag{15}$$

This can be extend trivially to a bound on the error in the signal space:

$$\|\mathbf{y} - \tilde{\mathbf{y}}\|_2^2 = \|\mathbf{\Phi s} - \mathbf{\Phi}\tilde{\mathbf{s}}\|_2^2 \leq C2(\log N_x)\sigma^p c^{1-\frac{p}{2}}\|\mathbf{\Phi}\|_2^2. \tag{16}$$

3.3 Random mixing conditions

In source separation, the mixing system should be considered independently from the dictionary $\mathbf{\Phi}$ in which the signal has a sparse representation. The theorems above are based on $\delta_m(\mathbf{M\Phi})$, which is required to be small. In this and the next section we derive new results that give insight into this quantity by considering $\mathbf{M}$ and $\mathbf{\Phi}$ separately.

The first theorem is a slight modification from [10][4] [5]:

Theorem 3 *Assume that* $\mathbf{M} \in \mathbb{R}^{N_x \times N_y}$ *is a random matrix with columns drawn uniformly from the unit sphere and let* $\mathbf{\Phi} \in \mathbb{R}^{N_y \times N_s}$ *have restricted isometry* $\delta_m(\mathbf{\Phi}) < 1$*, then there exists a constant* c*, such that for* $m \leq cN_x \log(N_s/m)$*:*

$$(1 - \delta_{P_m}(\mathbf{M})) \leq \frac{\|\mathbf{M\Phi}_\Gamma \mathbf{s}\|_2^2}{\|\mathbf{\Phi s}\|_2^2} \leq (1 + \delta_{P_m}(\mathbf{M})), \tag{17}$$

[4] Note, that there are a range of other distributions for which this theorem would hold, see [10] for details.

[5] Since the first submission of this manuscript we became aware of the paper [11], which contains very similar results.

and

$$(1-\delta_m(\mathbf{\Phi}))(1-\delta_{P_m}(\mathbf{M})) \le \frac{\|\mathbf{M\Phi}_\Gamma \mathbf{s}\|_2^2}{\|\mathbf{s}\|_2^2} \le (1+\delta_{P_m}(\mathbf{M}))(1+\delta_m(\mathbf{\Phi})), \quad (18)$$

holds with probability

$$\ge 1 - 2(eN_s/m)^m (12/\delta_{P_m})^m e^{-\frac{N_x}{2}(\delta_{P_m}^2/8-\delta_m^3/24)}. \quad (19)$$

Proof (Outline). The proof that equation (17) holds is similar to the proof given in [10] with the only difference that Theorem 5.1 in [10] can be shown to hold for any m dimensional subspace, and where N in theorem 5.2 in [10] can be replaced by N_s. The restricted isometry in equation (18) then follows by bounding $||\mathbf{\Phi s}||_2^2$ from above and below using the restricted isometry $(1-\delta_m(\mathbf{\Phi})) \le ||\mathbf{\Phi s}||_2^2 \le (1+\delta_m(\mathbf{\Phi}))$.

Therefore, for any dictionary $\mathbf{\Phi}$ with $\delta_m(\mathbf{\Phi}) < 1$ and for $\mathbf{M}$ sampled uniformly from the unit sphere, $\delta_m(\mathbf{M\Phi}) \le \delta_m(\mathbf{\Phi}) + \delta_{P_m}(M) + \delta_m(\mathbf{\Phi})\delta_{P_m}(M)$ with high probability, whenever $m \le CN_x/\log(N_s/N_x)$.

3.4 Non-random mixing matrix conditions

Unfortunately, theorem 3 assumes randomly generated mixing systems, which is rather restrictive. We therefore derive conditions that relate the measurement matrix M, the dictionary $\mathbf{\Phi}$ and $\delta_m(\mathbf{M\Phi})$.

To bound $\delta_m(\mathbf{M\Phi})$ we define the (to our knowledge novel) concept of M-coherence:

$$\mu_{\mathbf{M}}(\mathbf{\Phi}) = \max_{i,j:i\neq j} |\phi_i^T \mathbf{M}^T \mathbf{M} \phi_j|. \quad (20)$$

This quantity measures the coherence in the dictionary as 'seen through' the measuring matrix. We also need the quantities:

$$a_{min} = \min_i \|\mathbf{M}\phi_i\|_2^2 \text{ and } a_{max} = \max_i \|\mathbf{M}\phi_i\|_2^2, \quad (21)$$

which measure how much the measuring matrix can deform elements of the dictionary. We assume that $a_{min} \ge m\mu_{\mathbf{M}}(\mathbf{\Phi})$, then by the Gersgorin disk theorem for the eigenvalues of $\mathbf{\Phi}_\Gamma \mathbf{M}^T \mathbf{M} \mathbf{\Phi}_\Gamma$, we find that all squared singular values σ^2 of the matrix $\mathbf{M\Phi}_\Gamma$ with $|\Gamma| \le m$ are bounded by:

$$a_{min} - m\mu_{\mathbf{M}}(\mathbf{\Phi}) \le \sigma^2 \le a_{max} + m\mu_{\mathbf{M}}(\mathbf{\Phi}). \quad (22)$$

We therefore have the bound:

$$a_{min} - m\mu_{\mathbf{M}}(\mathbf{\Phi}) \le \frac{\|\mathbf{M\Phi}_\Gamma \mathbf{s}\|_2^2}{\|\mathbf{s}\|_2^2} \le a_{max} + m\mu_{\mathbf{M}}(\mathbf{\Phi}). \quad (23)$$

Using $a = \max\{a_{\max} - 1, 1 - a_{min}\}$ we have[6] the bound on $\delta_m(\mathbf{M\Phi})$ of

$$\delta_m(\mathbf{M\Phi}) \le a + m\mu_{\mathbf{M}}(\mathbf{\Phi}), \quad (24)$$

which is in terms of quantities that are easy to determine for a given dictionary $\mathbf{\Phi}$ and measurement matrix $\mathbf{M}$.

[6] Note that we have the bound $\|\mathbf{M}\|_2 \ge a_{max} \ge a_{min} \ge 0$.

3.5 Sensitivity to errors in M

In most source separation applications the mixing system is not given a priori and has to be estimated. This leads to the question of robustness of the method to errors in the estimation of the measuring matrix $\mathbf{M}$.

Assume we have an estimated mixing system $\tilde{\mathbf{M}} = \mathbf{M} + \mathbf{N}$ and estimated sources $\tilde{\mathbf{s}}$ such that $\tilde{\mathbf{x}} = \tilde{\mathbf{M}}\mathbf{\Phi}\tilde{\mathbf{s}}$ and such that $\tilde{\mathbf{s}}$ is supported on $\tilde{m}$ elements. Also assume that $\mathbf{x} = \mathbf{M}\mathbf{\Phi}\mathbf{s}$ is the true generating system with $\mathbf{s}$ supported on m elements. Further assume that $\|\tilde{\mathbf{x}} - \mathbf{x}\|_2 \leq 2\epsilon$.

If $\delta_{m+\tilde{m}}(\tilde{\mathbf{M}}\mathbf{\Phi}) < 1$, then we have the bound:

$$\|\mathbf{y} - \tilde{\mathbf{y}}\|_2 \leq \frac{\|\tilde{\mathbf{M}}\mathbf{\Phi}\mathbf{s} - \tilde{\mathbf{M}}\mathbf{\Phi}\tilde{\mathbf{s}}\|_2 \|\mathbf{\Phi}\|_2}{\sqrt{1 - \delta_{m+\tilde{m}}(\tilde{\mathbf{M}}\mathbf{\Phi})}}. \tag{25}$$

Replacing $\tilde{\mathbf{M}}\mathbf{\Phi}\mathbf{s}$ with $\mathbf{M}\mathbf{\Phi}\mathbf{s}+\mathbf{N}\mathbf{\Phi}\mathbf{s}$ and using $\|\mathbf{M}\mathbf{\Phi}\mathbf{s}-\tilde{\mathbf{M}}\mathbf{\Phi}\tilde{\mathbf{s}}\| \leq 2\epsilon$ together with the triangle inequality, we get the bound:

$$\|\mathbf{y} - \tilde{\mathbf{y}}\|_2 \leq \frac{2\epsilon + \|\mathbf{N}\mathbf{y}\|_2 \|\mathbf{\Phi}\|_2}{\sqrt{1 - \delta_{m+\tilde{m}}(\tilde{\mathbf{M}}\mathbf{\Phi})}} \leq \frac{2\epsilon + \|\mathbf{N}\|_2 \|\mathbf{\Phi}\|_2 \|\mathbf{y}\|_2}{\sqrt{1 - \delta_{m+\tilde{m}}(\tilde{\mathbf{M}}\mathbf{\Phi})}}. \tag{26}$$

4 Discussion and Conclusion

Underdetermined mixtures are a form of compressive sampling. Source separation is therefore equivalent to the decoding problem faced in compressive sampling. This equivalence opens up many new lines of enquiry, both in compressive sampling and in source separation.

On the one hand, as done in this paper, results form compressive sampling shed new insight into the source separation problem. For example, theorems 1 and 2 state that for signals with a sparse underlying representation, whether exact, or with decaying coefficients, convex optimisation techniques can be used to recover or approximate the original signal from a lower dimensional observation. Furthermore, results from compressive sampling give bounds on the estimation error for sources that have a sparse representation with decaying coefficients. The error is a function of this coefficient decay and properties of the matrix mapping the sparse representation into the mixed domain. In the source separation literature, linear programming techniques have been a common approach [3] [4] [7] and the new theory gives additional justification for the application of these techniques and, what is more, provides estimation bounds for certain problems.

The main novel contribution of this paper was an extension of recent results from compressive sampling to allow for more general, possibly over-complete dictionaries for the sparse representation. In source separation, the mixing matrix is in general unrelated to the dictionary and the main contribution of this paper was to derive conditions on the dictionary, the mixing system and their interaction that allow the application of standard compressive sampling results to the more general source separation problem. In particular we have disentangled the dictionary and the measurement matrix and could shown that for randomly generated mixing systems, the required conditions hold with high probability. For more general mixing systems, we have presented bounds on this condition, which are functions of simple to establish properties of the mixing system, the overcomplete dictionary and their interaction. If the mixing system has to be estimated as in many source separation settings, errors in this estimate will influence the estimates of the sources. The theory in subsection 3.5 gives bounds on this error.

Not only does source separation benefit from progress made in compressive sampling, compressive sampling has also much to learn from the extensive work done on source separation. For example, in source separation, the mixing system is not known in general and has to be estimated together with the sources. Many different estimation techniques have therefore been developed in the source separation community able to estimate the mixing system. This suggests an extension of compressive sampling to *blind compressive sampling* (BCS). Different scenarios seem possible depending on the application and, in our notation, either $\mathbf{\Phi}$ or $\mathbf{M}$ (or both) might be unknown or known only approximately. Preliminary work in this direction has shown encouraging first results and more formal studies are currently undertaken.

References

1. Donoho, D.: Compressed sensing. IEEE Trans. on Information Theory **52**(4) (2006) 1289–1306
2. Candès, E.: Compressive sampling. In: Proceedings of the International Congress of Mathematics, Madrid, Spain (2006)
3. Lewicki, M.S., Sejnowski, T.J.: Learning overcomplete representations. Neural Computation (12) (2000) 337–365
4. Zibulevsky, M., Bofill, P.: Underdetermined blind source separation using sparse representations. Signal Processing **81**(11) (2001) 2353–2362
5. Davies, M., Mitianoudis, N.: A simple mixture model for sparse overcomplete ICA. IEE Proc.-Vision, Image and Signal Processing **151**(1) (August 2004) 35–43
6. Fevotte, C., Godsill, S.: A bayesian approach for blind separation of sparse sources. IEEE Transactions on Speech and Audio Processing **PP**(99) (2005) 1–15
7. Zibulevsky, M., Pearlmutter, B.A.: Blind source separation by sparse decomposition in a signal dictionary. Neural Computation **13**(4) (2001) 863–882
8. Candes, E., Tao, T.: The dantzig selector: statistical estimation when p is larger than n. manuscript (2005)
9. Georgiev, P., Theis, F., Cichocki, A.: Sparse component analysis and blind source separation of underdetermined mixtures. IEEE Transactions on Neural Networks **16**(4) (2005) 992–996
10. Baraniuk, R., Davenport, M., De Vore, R., Wakin, M.: The Johnson-Lindenstrauss lemma meets compressed sensing. (2006)
11. Rauhut, H., Schnass, K., Vandergheynst, P.: Compressed sensing and redundant dictionaries. submitted to IEEE Transactions on Information Theory (2007)