

Independent Subspace Analysis Is Unique, Given Irreducibility

Harold W. Gutch and Fabian J. Theis

Max Planck Institute for Dynamics and Self-Organization, 37073 Göttingen, Germany
harold.gutch@ds.mpg.de, fabian@theis.name

Abstract. Independent Subspace Analysis (ISA) is a generalization of ICA. It tries to find a basis in which a given random vector can be decomposed into groups of mutually independent random vectors. Since the first introduction of ISA, various algorithms to solve this problem have been introduced, however a general proof of the uniqueness of ISA decompositions remained an open question. In this contribution we address this question and sketch a proof for the separability of ISA. The key condition for separability is to require the subspaces to be not further decomposable (irreducible). Based on a decomposition into irreducible components, we formulate a general model for ISA without restrictions on the group sizes. The validity of the uniqueness result is illustrated on a toy example. Moreover, an extension of ISA to subspace extraction is introduced and its indeterminacies are discussed.

With the increasing popularity of Independent Component Analysis, people started to get interested in extensions. Cardoso [2] was the first to formulate an extension denoted here as Independent Subspace Analysis. The general idea is that for a given observation \mathbf{X} we try to find an invertible matrix \mathbf{W} such that $\mathbf{W}\mathbf{X} = (\mathbf{S}_1^T, \dots, \mathbf{S}_k^T)^T$ with mutually independent random vectors \mathbf{S}_i . If all \mathbf{S}_i are one-dimensional, this is ICA, and we have the well-known separability results of ICA [3]. However without dimensionality restrictions, if mutual independence of the vectors \mathbf{S}_i is the only restriction imposed on \mathbf{W} , ISA cannot produce meaningful results: if \mathbf{W} simply is the identity and $k = 1$, then $\mathbf{S}_1 = \mathbf{X}$, which is independent of the (non-existing) rest. So, further restrictions are required for a meaningful model. A common approach is to fix the group size in advance, see [5] for a short review of ISA models. Here, we propose a more general concept based on [5], namely irreducibility of the recovered sources \mathbf{S}_i that is the requirement that any \mathbf{S}_i cannot be further decomposed. Our main contribution is a sound proof for the separability of this model together with a confirming simulation, thereby giving the details for the proposed ISA model from [5].

The manuscript is organized as follows. In the next section, we motivate the existence of such a separability result by studying a toy example. Then we give the sketch of the proof, and finally extend it to blind subspace extraction.

1 Motivation

Usually ISA is seen as a byproduct of ICA algorithms, which are assumed to decompose signals into components ‘as independent as possible’; the components are then simply sorted to give a decomposition into higher-dimensional subspaces. However this approach is not as straight-forward as it might seem, as, strictly speaking, if ICA is performed on a data set that cannot be completely decomposed into one-dimensional independent components, we are applying ICA to a data set that does not follow the ICA model and have no theoretical results predicting the behavior of ICA algorithms. Here we present some simulations, which give a hint that indeed ISA might not be so unproblematic.

We generated a toy data set consisting of two independent sources, each of which were not further decomposable. The first data set consisted of a wireframe model of a 3-dimensional cube, the second data set was created from a solid 2-dimensional circle, see figure 1. We uniformly picked $N = 10.000$ samples and mixed them in batch runs by applying $M = 200.000$ uniformly sampled orthogonal matrices. The 200.000 matrices were sampled, by choosing random matrices \mathbf{B} with entries normally sampled with mean 0 and variance 1, which then were symmetrically orthogonalized by $\mathbf{A} = (\mathbf{B}\mathbf{B}^T)^{-0.5}\mathbf{B}$. A mixture with an orthogonal matrix deviating from the block-structure should also deviate from independence within the blocks. As an ad-hoc measure for dependence within the blocks, we used the forth-order cumulant tensor:

$$\delta_{\mathbf{D}}(\mathbf{X}) = \sum_{i=1}^3 \sum_{j=4}^5 \sum_{k=1}^5 \sum_{l=1}^5 \text{cum}^2(X_i, X_j, X_k, X_l) .$$

This is motivated by the well-known and in ICA often used fact that the crosscumulant tensor is zero, i.e. $\text{cum}^2(\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_*, \mathbf{Y}_*) = 0$, if \mathbf{Y}_1 and \mathbf{Y}_2 are independent. We measured the deviation of our mixing matrices from block-structure by simply taking the Frobenius-norm of the off-block-diagonal blocks:

$$\text{off}(\mathbf{A}) := \sum_{i=1}^3 \sum_{j=4}^5 (a_{ij}^2 + a_{ji}^2) .$$

If ISA actually guarantees a unique block-structure in fourth order, we should get a dependence of 0 only if the mixing matrix itself is block-diagonal that is if $\text{off} \mathbf{A} = 0$. However, due to sampling errors, this is of course never reached, so we estimate the minima of $\delta_{\mathbf{D}}$. Figure 2 shows the relation of $\text{off} \mathbf{A}$ and $\delta_{\mathbf{D}}(\mathbf{A}\mathbf{S})$, and here we observe not only the expected minimum at $\text{off} \mathbf{A} = 0$, but two additional minima at $\text{off} \mathbf{A} = 2$ and $\text{off} \mathbf{A} = 4$. In order to take a closer look at these three points, we chose three matrices \mathbf{A}_0 , \mathbf{A}_2 and \mathbf{A}_4 , corresponding to the three local minima of the plot in Fig. 2. Starting with these matrices, we performed in their neighborhood a search for matrices with a lower model deviation. Again we sampled random orthogonal matrices, but this time biased them to be close to the identity matrix, as we wanted to search locally. We therefore again orthogonalized matrices as above, however chose the matrices \mathbf{B}

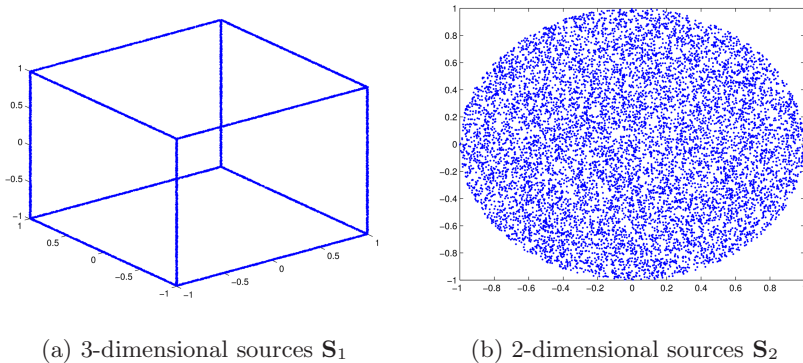


Fig. 1. Toy data set

to be not arbitrarily normally sampled, but took matrices whose entries were normally sampled with mean 0 and variance v , which we then added to the identity matrix, modifying \mathbf{A}_0 , \mathbf{A}_2 and \mathbf{A}_4 in every step only if it would perform a better block-independence. We evaluated this for $v = 0.1$, $v = 0.01$ and $v = 0.001$, each time running for 20,000 steps. The result of this is plotted in Fig. 3, and we indeed observe considerably better block-independence in the order of 1.5 magnitudes in the neighborhood of \mathbf{A}_0 than in the neighborhoods of \mathbf{A}_2 and \mathbf{A}_4 . While the three local minima found by random sampling show only small difference ($\delta_{\mathbb{D}}(\mathbf{A}_0) = 0.0135$, $\delta_{\mathbb{D}}(\mathbf{A}_2) = 0.0107$, $\delta_{\mathbb{D}}(\mathbf{A}_4) = 0.0285$), local searches show up better minima for all three areas ($\delta_{\mathbb{D}}(\mathbf{A}_0) = 0.0002$, $\delta_{\mathbb{D}}(\mathbf{A}_2) = 0.0055$, $\delta_{\mathbb{D}}(\mathbf{A}_4) = 0.0053$), especially the area around off $\mathbf{A} = 0$. As a side note, the final matrices \mathbf{A}_2 and \mathbf{A}_4 correspond to the product of a block-diagonal matrix and a permutation matrices where one, respectively two indices in each of the two off-diagonal blocks are non-zero.

This shows us that while we observe local minima of our block-dependency measure on our data set, a closer inspection reveals that these minima are of different quality and we actually have only a single global minimum. We conclude that separability of ISA indeed should hold.

2 Uniqueness of ISA

In this section we present the proof of uniqueness of ISA. After explaining the notion of *irreducibility* of a random vector, we show why this idea is essential for the separability of ISA.

2.1 The ICA Model

Let us quickly repeat a few facts about ICA. The linear, noiseless ICA model can be described by the equation $\mathbf{X} = \mathbf{A}\mathbf{S}$, where $\mathbf{S} = (S_1, \dots, S_n)^T$ denotes a random vector with mutually independent components S_i (sources) and an

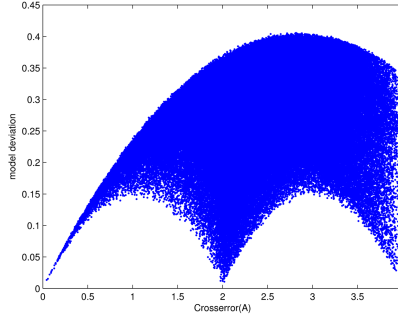


Fig. 2. Relation between block-crosserror and block-independence. Note the two additional minima at off $\mathbf{A} = 2$ and off $\mathbf{A} = 4$.

invertible mixing matrix \mathbf{A} . The task of ICA is the recovery of \mathbf{S} , given only the observations \mathbf{X} . This is obviously only possible up to the indeterminacies scaling and permutation, and it is well-known that recovery is possible up to exactly these permutations if \mathbf{S} is square-integrable and contains at most one Gaussian component [3, 4].

2.2 The ISA Model

Loosening the requirement of mutual independence of the sources naturally brings up the idea of describing ISA through the same equation $\mathbf{X} = \mathbf{A}\mathbf{S}$, where now $\mathbf{S} = (\mathbf{S}_1^T, \mathbf{S}_2^T, \dots, \mathbf{S}_n^T)^T$ with mutually independent random vectors \mathbf{S}_i , however this time dependencies within the multidimensional \mathbf{S}_i are allowed. Obvious indeterminacies of such a model are invertible linear transforms within the subspaces \mathbf{S}_i (which can be seen as a generalization of scaling to higher dimensions) and permutations of subspaces of the same size (which, again, is the higher dimensional generalization of the regular permutation seen in ICA). However this model is not complete, since for any observation \mathbf{X} a decomposition into mutually independent subspaces where dependencies within the subspaces are allowed is given simply by \mathbf{X} itself. Realizing this naturally brings up the requirement of \mathbf{S} to be ‘as independent as possible’. This is formally described by the following definition.

Definition 1. *A random vector \mathbf{S} is said to be irreducible if it contains no lower-dimensional independent component. An invertible matrix \mathbf{W} is called a (general) independent subspace analysis of \mathbf{X} if $\mathbf{W}\mathbf{X} = (\mathbf{S}_1^T, \dots, \mathbf{S}_k^T)^T$ with mutually independent, irreducible random vectors \mathbf{S}_i . Then $(\mathbf{S}_1^T, \dots, \mathbf{S}_k^T)$ is called an irreducible decomposition of \mathbf{X} .*

Irreducibility is a key property in uniqueness of ISA and indeed, if we additionally assume irreducibility, we can show that this essentially allows for separability of ISA up to the above mentioned indeterminacies of higher dimensional scaling and permutation of subspaces of the same size.

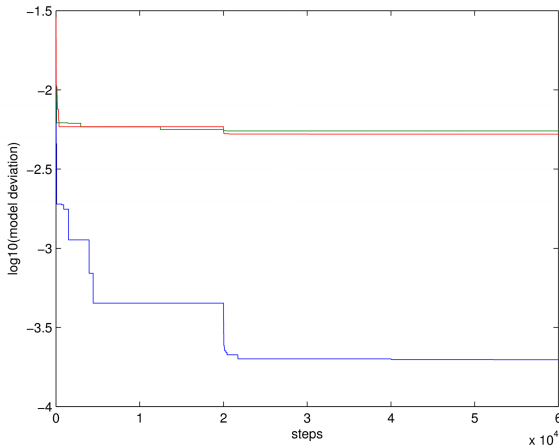


Fig. 3. Search for local minima around off $\mathbf{A} = 0$ (lower graph) and off $\mathbf{A} = 2$ respectively off $\mathbf{A} = 4$ (upper two graphs)

2.3 Uniqueness of ISA

We will now prove uniqueness of Independent Subspace Analysis under the additional assumption of no independent Gaussian components. Indeed, any orthogonal transformation of two decorrelated (and hence independent) Gaussians is again independent, so for such random vectors clearly such a strong identification result would not be possible.

Theorem 1. *Given a random vector \mathbf{X} with existing covariance and no Gaussian independent component, then an ISA of \mathbf{X} exists and is unique except for scaling and permutation.*

Existence holds trivially, but uniqueness is not obvious. Defining the equivalence relation \sim on random vectors as $\mathbf{X} \sim \mathbf{Y} :\Leftrightarrow \mathbf{X} = \mathbf{A}\mathbf{Y}$ for some $\mathbf{A} \in Gl(n)$, we are easily able to show uniqueness given the following lemma:

Lemma 1. *Let $\mathbf{S} = (\mathbf{S}_1^T, \dots, \mathbf{S}_N^T)^T$ be a square-integrable decomposition of \mathbf{S} into irreducible, mutually independent components \mathbf{S}_i where no \mathbf{S}_i is a one-dimensional Gaussian. If $(\mathbf{X}_1^T, \mathbf{X}_2^T)^T$ is an independent decomposition of \mathbf{S} , then there is some permutation π of $\{1, \dots, N\}$ such that $\mathbf{X}_1 \sim (\mathbf{S}_{\pi(1)}^T, \dots, \mathbf{S}_{\pi(l)}^T)^T$ and $\mathbf{X}_2 \sim (\mathbf{S}_{\pi(l+1)}^T, \dots, \mathbf{S}_{\pi(N)}^T)^T$ for some l .*

So, given an irreducible decomposition of a random variable \mathbf{S} with no independent Gaussian components, *any* decomposition of it into independent (not necessarily irreducible) components ‘splits along the irreducible components’.

Using this lemma, Theorem 1 is easy to show: Given two irreducible decompositions $(\mathbf{X}_1^T, \dots, \mathbf{X}_N^T)^T$ and $(\mathbf{S}_1^T, \dots, \mathbf{S}_M^T)^T$, we search for the smallest irreducible component appearing, which we may assume to be \mathbf{X}_1 . We then group

$(\mathbf{X}_2^T, \dots, \mathbf{X}_N^T)^T$ into a (larger) random vector. As this *independent* decomposition splits along the irreducible components \mathbf{S}_i and for all j , $\dim(\mathbf{X}_1) \leq \dim(\mathbf{S}_j)$, \mathbf{X}_1 is identical to one of the \mathbf{S}_j . We may remove both of these and go on iteratively, thus proving the theorem.

The more complicated part is the proof of Lemma 1, and due to space restrictions we can only sketch the proof.

Before starting, we note that due to the assumption of existing covariance, we may whiten both \mathbf{X} and \mathbf{S} , in which case it is easy to observe that \mathbf{A} is orthogonal. For notational reasons, we will split up the mixing matrix \mathbf{A} into submatrices, the sizes of which are according to the sizes of \mathbf{S}_i and \mathbf{X}_j :

$$\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_{11} & \dots & \mathbf{A}_{1N} \\ \mathbf{A}_{21} & \dots & \mathbf{A}_{2N} \end{pmatrix} \begin{pmatrix} \mathbf{S}_1 \\ \vdots \\ \mathbf{S}_N \end{pmatrix} \quad (1)$$

so $\mathbf{X}_i = \sum_{k=1}^N \mathbf{A}_{ik} \mathbf{S}_k$. We now claim that in every pair $\{\mathbf{A}_{1j}, \mathbf{A}_{2j}\}$ one of the two matrices is zero.

We fix $k = k_0$ and show this claim for k_0 . Let us assume the converse, that is that both $\text{rank}(\mathbf{A}_{1k_0}) \neq 0$ and $\text{rank}(\mathbf{A}_{2k_0}) \neq 0$. As \mathbf{A} has full rank, $\text{rank}(\mathbf{A}_{1k_0}) + \text{rank}(\mathbf{A}_{2k_0}) \geq \dim(\mathbf{S}_{k_0}) =: D$. This leaves us with two cases to handle, $\text{rank}(\mathbf{A}_{1k_0}) + \text{rank}(\mathbf{A}_{2k_0}) = D$ and $\text{rank}(\mathbf{A}_{1k_0}) + \text{rank}(\mathbf{A}_{2k_0}) > D$. Let us first address the first case and show that this contradicts the irreducibility of \mathbf{S}_{k_0} .

Lemma 2. *Assume*

$$\mathbf{S} = (\mathbf{A}_1 | \mathbf{A}_2) \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$$

with independent random vectors \mathbf{X}_1 and \mathbf{X}_2 and $\mathbf{A}_1, \mathbf{A}_2$ such that $\text{rank}(\mathbf{A}_1^T) + \text{rank}(\mathbf{A}_2^T) = \dim(\mathbf{S})$ and $\text{rank}(\mathbf{A}_1 | \mathbf{A}_2) = \dim(\mathbf{S})$. Then \mathbf{S} is reducible.

Proof. Let $D := \dim(\mathbf{S})$ and $d := \dim(\ker(\mathbf{A}_1^T))$. Then $\dim(\ker(\mathbf{A}_2^T)) = D - d$, and we can find a linearly independent set $\{\mathbf{v}_1, \dots, \mathbf{v}_d\}$ such that $\mathbf{v}_i^T \mathbf{A}_1 = 0$ for any $1 \leq i \leq d$, and similarly a linearly independent set $\{\mathbf{v}_{d+1}, \dots, \mathbf{v}_D\}$ such that $\mathbf{v}_j^T \mathbf{A}_2 = 0$ for any $d+1 \leq j \leq D$. These two sets are guaranteed to be disjoint, as $\text{rank}(\mathbf{A}_1 | \mathbf{A}_2) = \dim(\mathbf{S})$. Using these vectors, we define

$$\mathbf{T} := \begin{pmatrix} \mathbf{v}_1^T \\ \vdots \\ \mathbf{v}_D^T \end{pmatrix}.$$

Then

$$\mathbf{TS} = (\mathbf{TA}_1 | \mathbf{TA}_2) \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{T}_1 & | & 0 \\ 0 & | & \mathbf{T}_2 \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix} = \begin{pmatrix} \mathbf{T}_1 \mathbf{X}_1 \\ \mathbf{T}_2 \mathbf{X}_2 \end{pmatrix}$$

with some full rank matrices \mathbf{T}_1 and \mathbf{T}_2 . It follows that \mathbf{S} is reducible, as \mathbf{X}_1 and \mathbf{X}_2 are independent and \mathbf{T} is invertible. \square

The other case, $\text{rank}(\mathbf{A}_{1k_0}) + \text{rank}(\mathbf{A}_{2k_0}) > D$ is harder to prove and follows some of the ideas presented in [4].

Lemma 3. *Given (1), if there is some $1 \leq k_0 \leq N$ such that $\text{rank}(\mathbf{A}_{1k_0}) + \text{rank}(\mathbf{A}_{2k_0}) > \dim(\mathbf{S}_{k_0})$, then \mathbf{S}_{k_0} contains an irreducible Gaussian component.*

This concludes the proof of Theorem 1.

2.4 Dealing with Gaussians

The section above explicitly excluded independent Gaussian components in order to avoid additional indeterminacies. Recently, a general decomposition model dealing with Gaussians was proposed in the form of the so-called *non-Gaussian component analysis (NGCA)* [1]. It tries to detect a whole non-Gaussian subspace within the data, and no assumption of independence within the subspace is made. More precisely, given a random vector \mathbf{X} , a factorization $\mathbf{X} = \mathbf{A}\mathbf{S}$ with an invertible matrix \mathbf{A} , $\mathbf{S} = (\mathbf{S}_N, \mathbf{S}_G)$ and \mathbf{S}_N a square-integrable m -dimensional random vector is called an m -decomposition of \mathbf{X} if \mathbf{S}_N and \mathbf{S}_G are stochastically independent and \mathbf{S}_G is Gaussian. In this case, \mathbf{X} is said to be *m -decomposable* and \mathbf{X} is denoted to be *minimally n -decomposable* if \mathbf{X} is not $(n-1)$ -decomposable. According to our previous notation, \mathbf{S}_N and \mathbf{S}_G are independent components of \mathbf{X} . It has been shown that the subspaces of such decompositions are unique [6]:

Theorem 2. *The mixing matrix \mathbf{A} of a minimal decomposition is unique except for transformations in each of the two subspaces.*

Moreover, explicit algorithms can be constructed for identifying the subspaces [6]. This result enables us to generalize Theorem 1 and to get a general decomposition theorem, which characterizes solutions of ISA.

Theorem 3. *Given a random vector \mathbf{X} with existing covariance, an ISA of \mathbf{X} exists and is unique except for permutation of components of the same dimension and invertible transformations within each independent component and within the Gaussian part.*

Proof. Existence is obvious. Uniqueness follows after first applying Theorem 2 to \mathbf{X} and then Theorem 1 to the non-Gaussian part. \square

3 Independent Subspace Extraction

Having shown uniqueness of the decomposition, we are able to introduce Independent (Irreducible) Subspace Extraction, which separates independent (irreducible) subspaces out of the random vector.

Definition 2. *A pseudo-invertible $(n \times m)$ matrix \mathbf{W} is said to be an Independent Subspace Extraction of an m -dimensional random vector \mathbf{X} , if $\mathbf{W}\mathbf{X}$ is an independent component of \mathbf{X} . If $\mathbf{W}\mathbf{X}$ even is irreducible, then \mathbf{W} is called an Irreducible Subspace Extraction of \mathbf{X} .*

This could lead to a wider variety of algorithms like deflationary approaches which are already common in standard ICA. The interesting aspect here is that we only strive to extract a single component, so Independent (Irreducible) Subspace Extraction could prove to be simpler to handle algorithmically than a complete Independent Subspace Analysis, and thus play an important role in applications (such as dimension reduction) that need to extract only a single component or subspace.

4 Conclusion

Although Independent Subspace Analysis has become a common practice in the last few years, separability of it has not been fully shown. We presented examples that showed that ISA is not as unproblematic as it seems. Additionally we proved uniqueness – up to higher-dimensional generalizations of the indeterminacies of ICA – of ISA, given no independent Gaussians and showed how to combine this together with existing theoretical results on NGCA to a full ISA uniqueness result. Using these results, it is now possible to speak of *the* ISA of any given random vector. Moreover, theorem 3 now gives an complete characterization of decompositions of distributions into independent factors, which might prove to be a useful result in general statistics.

Now that uniqueness of ISA has been shown for the theoretical limit of perfect knowledge of the recordings, the next obvious step is the conversion to the real-world case, where only a finite number of samples of the observations are known. Here, a decomposition of the mixtures \mathbf{X} such that $\mathbf{X} = \mathbf{A}\mathbf{S}$ where $\mathbf{S} = (\mathbf{S}_1^T, \dots, \mathbf{S}_N^T)^T$ with irreducible (or merely independent) \mathbf{S}_i cannot be expected, as in this case we expect to always see some dependency due to sampling errors. Due to uniqueness of ISA in the asymptotic case, identification of the underlying sources should hold here too, given enough samples, but additional work is required to show this in the future.

References

1. Blanchard, G., Kawanabe, M., Sugiyama, M., Spokoiny, V., Müller, K.R.: In search of non-gaussian components of a high-dimensional distribution. *Journal of Machine Learning Research* 7, 247–282 (2006)
2. Cardoso, J.F.: Multidimensional independent component analysis. In: *Proc. of ICASSP '98*, Seattle (1998)
3. Comon, P.: Independent component analysis - a new concept? *Signal Processing* 36, 287–314 (1994)
4. Theis, F.J.: A new concept for separability problems in blind source separation. *Neural Computation* 16, 1827–1850 (2004)
5. Theis, F.J.: Towards a general independent subspace analysis. In: *Proc. NIPS 2006* (2007)
6. Theis, F.J., Kawanabe, M.: Uniqueness of non-gaussian subspace analysis. In: Rosca, J., Erdogmus, D., Príncipe, J.C., Haykin, S. (eds.) *ICA 2006*. LNCS, vol. 3889, pp. 917–925. Springer, Heidelberg (2006)