Infinite Sparse Factor Analysis and Infinite Independent Components Analysis

David Knowles and Zoubin Ghahramani

d.a.knowles.07@cantab.net, zoubin@eng.cam.ac.uk
Department of Engineering
University of Cambridge
CB2 1PZ, UK

Abstract. A nonparametric Bayesian extension of Independent Components Analysis (ICA) is proposed where observed data $\mathbf{Y}$ is modelled as a linear superposition, $\mathbf{G}$, of a potentially infinite number of hidden sources, $\mathbf{X}$. Whether a given source is active for a specific data point is specified by an infinite binary matrix, $\mathbf{Z}$. The resulting sparse representation allows increased data reduction compared to standard ICA. We define a prior on $\mathbf{Z}$ using the Indian Buffet Process (IBP). We describe four variants of the model, with Gaussian or Laplacian priors on $\mathbf{X}$ and the one or two-parameter IBPs. We demonstrate Bayesian inference under these models using a Markov Chain Monte Carlo (MCMC) algorithm on synthetic and gene expression data and compare to standard ICA algorithms.

1 Introduction

Independent Components Analysis (ICA) is a model which explains observed data, $\mathbf{y}_t$ (dimension D) in terms of a linear superposition of independent hidden sources, $\mathbf{x}_t$ (dimension K), so $\mathbf{y}_t = \mathbf{G}\mathbf{x}_t + \boldsymbol{\epsilon}_t$, where $\mathbf{G}$ is the mixing matrix and $\boldsymbol{\epsilon}_t$ is Gaussian noise. In the standard ICA model we assume $K = D$ and that there exists $\mathbf{W} = \mathbf{G}^{-1}$. We use FastICA [1], a widely used implementation, as a benchmark. The assumption $K = D$ may be invalid, so Reversible Jump MCMC [2] could be used to infer K. In this paper we propose a sparse implementation which allows a potentially infinite number of components and the choice of whether a hidden source is active for a data point, allowing increased data reduction for systems where sources are intermittently active. Although ICA is not a time-series model it has been used successfully on time-series data such as electroencephalograms [3]. It has also been applied to gene expression data [4], the application we choose for a demonstration.

2 The Model

We define a binary vector $\mathbf{z}_t$ which acts as a mask on $\mathbf{x}_t$. Element z_{kt} specifies whether hidden source k is active for data point t. Thus

$$\mathbf{Y} = \mathbf{G}(\mathbf{Z} \odot \mathbf{X}) + \mathbf{E} \tag{1}$$

where $\odot$ denotes element-wise multiplication and $\mathbf{X}$, $\mathbf{Y}$, $\mathbf{Z}$ and $\mathbf{E}$ are concatenated matrices of $\mathbf{x}_t$, $\mathbf{y}_t$, $\mathbf{z}_t$ and $\boldsymbol{\epsilon}_t$ respectively. We allow a potentially infinite number of hidden sources, so that $\mathbf{Z}$ has infinitely many rows, although only a finite number will have non-zero entries. We assume Gaussian noise with variance σ_ϵ^2, which is given an inverse Gamma prior $\mathcal{IG}\left(\sigma_\epsilon^2; a, b\right)$.

We define two variants based on the prior for x_{kt}: *infinite sparse Factor Analysis* (isFA) has a unit Gaussian prior; *infinite Independent Components Analysis* (iICA) has a Laplacian(1) prior. Other heavy tailed distributions are possible but are not explored here. Varying the variance is redundant because we infer the variance of the mixture weights. The prior on the elements of $\mathbf{G}$ is Gaussian with variance σ_G^2, which is given an inverse Gamma prior. We define the prior on $\mathbf{Z}$ using the Indian Buffet Process with parameter α (and later β) as described in Section 2.1 and in more detail in [5]. We place Gamma priors on α and β.

All four variants share

$$\boldsymbol{\epsilon}_t \sim \mathcal{N}\left(0, \sigma_\epsilon^2 \mathbf{I}\right) \qquad \sigma_\epsilon^2 \sim \mathcal{IG}\left(a, b\right) \tag{2}$$

$$\mathbf{g}_k \sim \mathcal{N}\left(0, \sigma_G^2\right) \qquad \sigma_G^2 \sim \mathcal{IG}\left(c, d\right) \tag{3}$$

$$\mathbf{Z} \sim \mathcal{IBP}(\alpha, \beta) \qquad \alpha \sim \mathcal{G}\left(e, f\right) \tag{4}$$

The differences between the variants are summarised here.

	$x_{kt} \sim \mathcal{N}(0,1)$	$x_{kt} \sim \mathcal{L}(1)$
$\beta = 1$	$isFA_1$	$iICA_1$
$\beta \sim \mathcal{G}(1,2)$	$isFA_2$	$iICA_2$

2.1 Defining a distribution on an infinite binary matrix

Start with a finite model. We derive our distribution on $\mathbf{Z}$ by defining a finite K model and taking the limit as $K \to \infty$. We then show how the infinite case corresponds to a simple stochastic process.

We assume that the probability of a source k being active is π_k, and that the sources are generated independently. We find

$$P(\mathbf{Z}|\boldsymbol{\pi}) = \prod_{k=1}^{K} \prod_{t=1}^{N} P(z_{kt}|\pi_k) = \prod_{k=1}^{K} \pi_k^{m_k} (1-\pi_k)^{N-m_k} \tag{5}$$

where N is the total number of data points and $m_k = \sum_{t=1}^{N} z_{kt}$ is the number of data points for which source k is active. We put a $\text{Beta}(\frac{\alpha}{K}, 1)$ prior on π_k, where α is the strength parameter. Due to the conjugacy between the binomial and beta distributions we are able to integrate out π to find

$$P(\mathbf{Z}) = \prod_{k=1}^{K} \frac{\frac{\alpha}{K}\Gamma(m_k + \frac{\alpha}{K})\Gamma(N - m_k + 1)}{\Gamma(N + 1 + \frac{\alpha}{K})} \tag{6}$$

Take the infinite limit. By defining a scheme to order the non-zero rows of $\mathbf{Z}$ (see [5]) we can take $K \to \infty$ and find

$$P(\mathbf{Z}) = \frac{\alpha^{K_+}}{\prod_{h>0} K_h!} \exp\{-\alpha H_N\} \prod_{k=1}^{K_+} \frac{(N-m_k)!(m_k-1)!}{N!} \tag{7}$$

where K_+ is the number of active features, $H_N = \sum_{j=1}^{N} \frac{1}{j}$ is the N-th harmonic number, and K_h is the number of rows whose entries correspond to the binary number h.

Go to an Indian Buffet. This distribution corresponds to a simple stochastic process, the Indian Buffet Process. Consider a buffet with a seemingly infinite number of dishes (hidden sources) arranged in a line. The first customer (data point) starts at the left and samples Poisson(α) dishes. The ith customer moves from left to right sampling dishes with probability $\frac{m_k}{i}$ where m_k is the number of customers to have previously sampled that dish. Having reached the end of the previously sampled dishes, he tries Poisson($\frac{\alpha}{i}$) new dishes.

If we apply the same ordering scheme to the matrix generated by this process as for the finite model, we recover the correct exchangeable distribution. Since the distribution is exchangeable with respect to the customers we find by considering the last customer that $P(z_{kt} = 1|\mathbf{z}_{-kt}) = \frac{m_{k,-t}}{N}$ where $m_{k,-t} = \sum_{s \neq t} z_{ks}$, which is used in sampling $\mathbf{Z}$. By exchangeability and considering the first customer, the number of active sources for a data point follows a Poisson(α) distribution, and the expected number of entries in $\mathbf{Z}$ is $N\alpha$. We also see that the number of active features, $K_+ = \sum_{t=1}^{N} \text{Poisson}(\frac{\alpha}{t}) = \text{Poisson}(\alpha H_N)$ which grows as $\alpha \log N$ for large N.

Two parameter generalisation. A problem with the one parameter IBP is that the number of features per object, α, and the total number of features, $N\alpha$, are both controlled by α and cannot vary independently. Under this model, we cannot tune how likely it is for features to be shared across objects. To overcome this restriction we follow [6], introducing β, a measure of the feature *repulsion*. The ith customer now samples dish k with probability $\frac{m_k}{\beta+i-1}$ and samples Poisson($\frac{\alpha\beta}{\beta+i-1}$) new dishes.

We find $P(z_{kt} = 1|\mathbf{z}_{-kt}, \beta) = \frac{m_{k,-t}}{\beta+N-1}$. The marginal probability of $\mathbf{Z}$ becomes

$$P(\mathbf{Z}|\alpha, \beta) = \frac{(\alpha\beta)^{K_+}}{\prod_{h>0} K_h!} \exp\{-\alpha H_N(\beta)\} \prod_{k=1}^{K_+} B(m_k, N - m_k + \beta) \tag{8}$$

where $H_N(\beta) = \sum_{j=1}^{N} \frac{\beta}{\beta+j-1}$.

3 Inference

Given the observed data $\mathbf{Y}$, we wish to infer the hidden sources $\mathbf{X}$, which sources are active $\mathbf{Z}$, the mixing matrix $\mathbf{G}$, and all hyperparameters. We use Gibbs sam-

pling, but with Metropolis-Hastings (MH) steps for β and sampling new features. We draw samples from the marginal distribution of the model parameters given the data by successively sampling the conditional distributions of each parameter in turn, given all other parameters.

Hidden sources. We sample each element of $\mathbf{X}$ for which $z_{kt} = 1$. We denote the k-th column of $\mathbf{G}$ by $\mathbf{g}_k$ and $\boldsymbol{\epsilon}_t|_{z_{kt}=0}$ by $\boldsymbol{\epsilon}_{-kt}$. For isFA we find the conditional distribution is a Gaussian:

$$P(x_{kt}|\mathbf{G}, \mathbf{x}_{-kt}, \mathbf{y}_t, \mathbf{z}_t) = \mathcal{N}\left(x_{kt}; \frac{\mathbf{g}_k^T \boldsymbol{\epsilon}_{-kt}}{\sigma_\epsilon^2 + \mathbf{g}_k^T \mathbf{g}_k}, \frac{\sigma_\epsilon^2}{\sigma_\epsilon^2 + \mathbf{g}_k^T \mathbf{g}_k}\right) \tag{9}$$

For iICA we find a piecewise Gaussian distribution, which it is possible to sample from analytically given the Gaussian c.d.f. function F

$$P(x_{kt}|\mathbf{G}, \mathbf{x}_{-kt}, \mathbf{y}_t, \mathbf{z}_t) = \begin{cases} \mathcal{N}\left(x_{kt}; \mu_-, \sigma^2\right) & x_{kt} > 0 \\ \mathcal{N}\left(x_{kt}; \mu_+, \sigma^2\right) & x_{kt} < 0 \end{cases} \tag{10}$$

$$\text{where } \mu_\pm = \frac{\mathbf{g}_k^T \boldsymbol{\epsilon}_{-kt} \pm \sigma_\epsilon^2}{\mathbf{g}_k^T \mathbf{g}_k} \text{ and } \sigma^2 = \frac{\sigma_\epsilon^2}{\mathbf{g}_k^T \mathbf{g}_k} \tag{11}$$

Active sources. To sample $\mathbf{Z}$ we first define the ratio of conditionals, r

$$r = \underbrace{\frac{P(\mathbf{y}_t|\mathbf{G}, \mathbf{x}_{-kt}, \mathbf{z}_{-kt}, z_{kt} = 1, \sigma_\epsilon^2)}{P(\mathbf{y}_t|\mathbf{G}, \mathbf{x}_{-kt}, \mathbf{z}_{-kt}, z_{kt} = 0, \sigma_\epsilon^2)}}_{r_l} \underbrace{\frac{P(z_{kt} = 1|\mathbf{z}_{-kt})}{P(z_{kt} = 0|\mathbf{z}_{-kt})}}_{r_p} \tag{12}$$

so that $P(z_{kt} = 1|\mathbf{G}, \mathbf{X}_{-kt}, \mathbf{Y}, \mathbf{Z}_{-kt}) = \frac{r}{r+1}$. From Section 2.1 the ratio of priors is $r_p = \frac{m_{k,-t}}{\beta + N - 1 - m_{k,-t}}$. To find $P(\mathbf{y}_t|\mathbf{G}, \mathbf{x}_{-kt}, \mathbf{z}_{-kt}, z_{kt} = 1)$ we must marginalise over all possible values of x_{kt}.

$$P(\mathbf{y}_t|\mathbf{G}, \mathbf{x}_{-kt}, \mathbf{z}_{-kt}, z_{kt} = 1) = \int P(\mathbf{y}_t|\mathbf{G}, \mathbf{x}_t, \mathbf{z}_{-kt}, z_{kt} = 1) P(x_{kt}) \mathrm{d}x_{kt} \tag{13}$$

For isFA, using Equation (9) and integrating we find $r_l = \sigma \exp\left\{\frac{\mu^2}{2\sigma^2}\right\}$. For iICA we use Equation (11) and integrate above and below 0 to find

$$r_l = \sigma\sqrt{\frac{\pi}{2}}\left[F(0; \mu_+, \sigma) \exp\left\{\frac{\mu_+^2}{2\sigma^2}\right\} + (1 - F(0; \mu_-, \sigma)) \exp\left\{\frac{\mu_-^2}{2\sigma^2}\right\}\right] \tag{14}$$

Creating new features. $\mathbf{Z}$ is a matrix with infinitely many rows, but only the non-zero rows can be held in memory. However, the zero rows still need to be taken into account. Let κ_t be the number of rows of $\mathbf{Z}$ which contain 1 only in column t, i.e. the number of features which are active only at time t. New features are proposed by sampling κ_t with a MH step. We propose a move

$\xi \to \xi^*$ with probability $J(\xi^*|\xi)$, following [7], we set to be equal to the prior on ξ^*. This move is accepted with probability $\min(1, r_{\xi\to\xi^*})$ where

$$r_{\xi\to\xi^*} = \frac{P(\xi^*|\text{rest})J(\xi|\xi^*)}{P(\xi^*|\text{rest})J(\xi^*|\xi)} = \frac{P(\text{rest}|\xi^*)P(\xi^*)P(\xi)}{P(\text{rest}|\xi)P(\xi)P(\xi^*)} = \frac{P(\text{rest}|\xi^*)}{P(\text{rest}|\xi)} \tag{15}$$

where *rest* denotes all other variables. By this choice $r_{\xi\to\xi^*}$ becomes the ratio of likelihoods. From the IBP the prior for κ_t is $P(\kappa_t|\alpha) = \text{Poisson}(\frac{\alpha\beta}{\beta+N-1})$.

For isFA we can integrate out $\mathbf{x}'_t$, the new elements of $\mathbf{x}_t$, but not $\mathbf{G}'$, the new columns of $\mathbf{G}$, so our proposal is $\xi = \{\mathbf{G}', \kappa_t\}$. We find $r_{\xi\to\xi^*} = |\boldsymbol{\Lambda}|^{-\frac{1}{2}} \exp\left(\frac{1}{2}\boldsymbol{\mu}^T\boldsymbol{\Lambda}\boldsymbol{\mu}\right)$ where $\boldsymbol{\Lambda} = \mathbf{I} + \frac{\mathbf{G}^{*T}\mathbf{G}^*}{\sigma_\epsilon^2}$ and $\boldsymbol{\Lambda}\boldsymbol{\mu} = \frac{1}{\sigma_\epsilon^2}\mathbf{G}^{*T}\boldsymbol{\epsilon}_t$.

For iICA marginalisation is not possible so $\xi = \{\mathbf{G}', \mathbf{x}'_t, \kappa_t\}$. From Equation (15) we find

$$r_{\xi\to\xi^*} = \exp\left\{-\frac{1}{2\sigma_\epsilon^2}\mathbf{x}_t'^T\mathbf{G}^{*T}(\mathbf{G}^*\mathbf{x}'_t - 2\boldsymbol{\epsilon}_t)\right\} \tag{16}$$

Mixture weights. We sample the columns $\mathbf{g}_k$ of $\mathbf{G}$. We denote the kth row of $(\mathbf{Z}\odot\mathbf{X})$ by $\mathbf{x}_k'^T$. We have $P(\mathbf{g}_k|\mathbf{G}_{-k}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}, \sigma_\epsilon^2, \sigma_G^2) \propto P(\mathbf{Y}|\mathbf{G}, \mathbf{X}, \mathbf{Z}, \sigma_\epsilon^2)P(\mathbf{g}_k|\sigma_G^2)$. The total likelihood function has exponent

$$-\frac{1}{2\sigma_\epsilon^2}\operatorname{tr}(\mathbf{E}^T\mathbf{E}) = -\frac{1}{2\sigma_\epsilon^2}((\mathbf{x}_k'^T\mathbf{x}'_k)(\mathbf{g}_k^T\mathbf{g}_k) - 2\mathbf{g}_k^T\mathbf{E}|_{\mathbf{g}_k=0}) + \text{const} \tag{17}$$

where $\mathbf{E} = \mathbf{Y} - \mathbf{G}(\mathbf{Z}\odot\mathbf{X})$. We thus find the conditional of $\mathbf{g}_k$ is $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ where $\boldsymbol{\mu} = \frac{\sigma_G^2}{\mathbf{x}_k'^T\mathbf{x}'_k\sigma_G^2+\sigma_\epsilon^2}\mathbf{E}|_{\mathbf{g}_k=0}\mathbf{x}'_k$ and $\boldsymbol{\Lambda} = \left(\frac{\mathbf{x}_k'^T\mathbf{x}'_k}{\sigma_\epsilon^2} + \frac{1}{\sigma_G^2}\right)\mathbf{I}_{D\times D}$.

Learning the noise level. We allow the model to learn the noise level σ_ϵ^2. Applying Bayes' rule we find

$$P(\sigma_\epsilon^2|\mathbf{E}, a, b) \propto P(\mathbf{E}|\sigma_\epsilon^2)P(\sigma_\epsilon^2|a, b) = \mathcal{IG}\left(\sigma_\epsilon^2; a + \frac{ND}{2}, \frac{b}{1 + \frac{b}{2}\operatorname{tr}(\mathbf{E}^T\mathbf{E})}\right) \tag{18}$$

Inferring the scale of the data. For sampling σ_G^2 the conditional prior on $\mathbf{G}$ acts as the likelihood term

$$P(\sigma_G^2|\mathbf{G}, c, d) \propto P(\mathbf{G}|\sigma_G^2)P(\sigma_G^2|c, d) = \mathcal{IG}\left(\sigma_G^2; c + \frac{DK}{2}, \frac{d}{1 + \frac{d}{2}\operatorname{tr}(\mathbf{G}^T\mathbf{G})}\right) \tag{19}$$

IBP parameters. We infer the IBP strength parameter α. The conditional prior on $\mathbf{Z}$, given by Equation (8), acts as the likelihood term

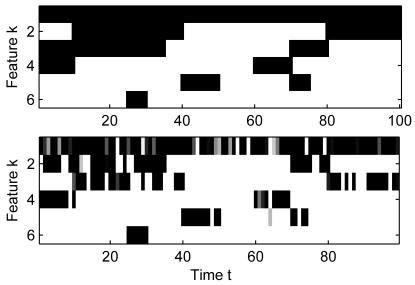
$$P(\alpha|\mathbf{Z}, \beta) \propto P(\mathbf{Z}|\alpha, \beta)P(\alpha) = \mathcal{G}\left(\alpha; K_+ + e, \frac{f}{1 + fH_N(\beta)}\right) \tag{20}$$

We sample β by a MH step with acceptance probability $\min(1, r_{\beta\to\beta^*})$. By Equation (15) setting $J(\beta^*|\beta) = P(\beta^*) = \mathcal{G}(1, 1)$, results in $r_{\beta\to\beta^*} = \frac{P(Z|\alpha,\beta^*)}{P(Z|\alpha,\beta)}$.

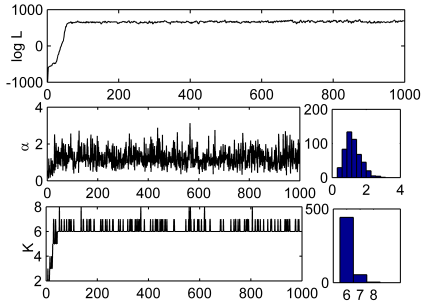
4 Results

4.1 Synthetic data

We ran all four variants and three FastICA variants (using the *pow3, tanh* and *gauss* non-linearities) on 30 sets of randomly generated data with $D = 7, K = 6, N = 200$, the $\mathbf{Z}$ matrix shown in Figure 1(a), and Gaussian or Laplacian source distributions. Figure 1 shows the average inferred $\mathbf{Z}$ matrix and algorithm convergence for a typical 1000 iteration ICA_1 run. $\mathbf{Z}$ is successfully recovered within an arbitrary ordering. The gaps in the inferred $\mathbf{Z}$ are a result of inferring $z_{kt} = 0$ where $x_{kt} = 0$.



(a) *Top:* True $\mathbf{Z}$. *Bottom:* Inferred $\mathbf{Z}$.



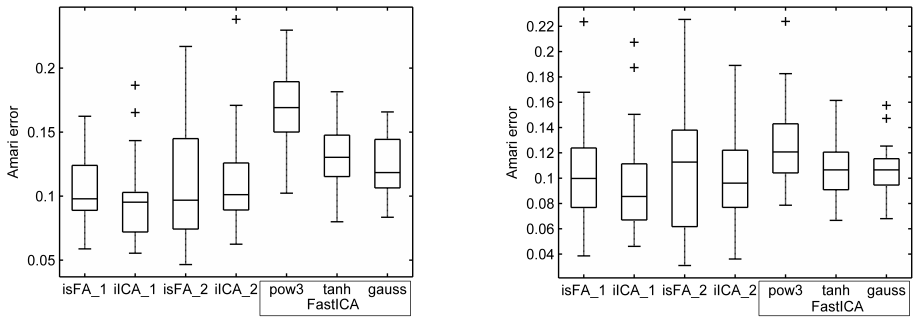
(b) Log likelihood, α and K^+ for duration of 1000 iteration run.

Fig. 1. True and inferred $\mathbf{Z}$ and algorithm convergence for typical iICA_1 run.

Boxplots of the Amari error [8] for each algorithm are shown in Figure 2. Figure 2(a) shows the results when the synthetic data has Gaussian source distributions. All four variants perform significantly better on the sparse synthetic data than any of the FastICA variants, but then we do not expect FastICA to recover Gaussian sources. Figure 2(b) shows the results when the synthetic data has Laplacian source distributions. As expected the FastICA performance is much improved because the sources are heavy tailed. However, isFA_1, iICA_1 and iICA_2 still perform better on average because they correctly model the sparse nature of the data. The performance of isFA_2 is severely effected by having the incorrect source model, suggesting the iICA variants may be more robust to deviations from the assumed source distribution. The two parameter IBP variants of both algorithms actually perform no better than the one parameter versions: $\beta = 1$ happens to be almost optimal for the synthetic $\mathbf{Z}$ used.

4.2 Gene expression data

We now apply our model to the microarray data from an ovarian cancer study [4], which represents the expression level of $N = 172$ genes (data points) across



(a) Gaussian sources. (b) Laplacian sources.

Fig. 2. Boxplots of Amari errors for 30 synthetic data sets with $D = 7, N = 6, N = 100$ analysed using each algorithm variant and FastICA.

$D = 17$ tissue samples (observed variables). The tissue samples are grouped into five tissue types: one healthy and four diseased. ICA was applied to this dataset in [4], where the term *gene signature* is used to describe the infered hidden sources. Some of the processes which regulate gene expression, such as DNA methylation, completely *silence* the gene, while others, such as transcription regulation, affect the level at which the gene is expressed. Thus our sparse model is highly valid for this system: $\mathbf{Z}$ represents which genes are silenced, and $\mathbf{X}$ represents the expression level of active genes.

Figure 4.2 shows the mean $\mathbf{G}$ matrix infered by iICA_1. Gene signature (hidden source) 1 is expressed across all the tissue samples, accounted for genes shared by all the samples. Signature 7 is specific to the pd-spa tissue type. This is consistent with that found in [4], with the same top 3 genes. Such tissue type dependent signatures could be used for observer independent classification. Signatures such as 5 which is differentially expressed across the pd-spa samples could help subclassify tissue types.

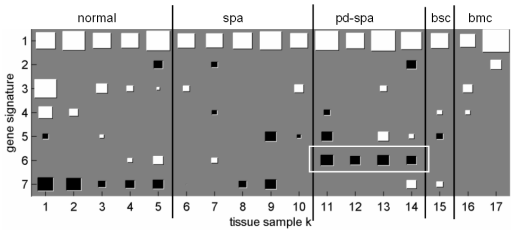


Fig. 3. Hinton diagram of $\mathbf{G}^T$: the expression level of each gene signature within each tissue sample.

5 Conclusions and Future work

In this paper we have defined the Infinite Sparse FA and Infinite ICA models using a distribution over the infinite binary matrix $\mathbf{Z}$ corresponding to the Indian Buffet Process. We have derived MCMC algorithms for each model to infer the parameters given observed data. These have been demonstrated on synthetic data, where the correct assumption about the hidden source distribution was shown to give optimal performance, and gene expression data, where the results were consistent with those using ICA. A MATLAB implementation of the algorithms will be made available at http://learning.eng.cam.ac.uk/zoubin/.

There are a number of directions in which this work can be extended. The recently developed stick breaking constructions for the IBP will allow a slice sampler to be derived for $\mathbf{Z}$ which should allow faster mixing than the MH step currently used in sampling new features. Faster partially deterministic algorithms would be useful for online learning in applications such as audio processing. The sparse nature of the model could have useful applications in data compression for storage or data reduction for further analysis.

References

1. Hyvärinen, A.: Fast and robust fixed-point algorithms for independent component analysis. IEEE Transactions on Neural Networks **10**(3) (1999) 626–634
2. Richardson, S., Green, P.J.: On bayesian analysis of mixtures with an unknown number of components. Journal of the Royal Statistical Society **59** (1997) 731–792
3. Makeig, S., Bell, A.J., Jung, T.P., Sejnowski, T.J.: Independent component analysis of electroencephalographic data. Advances in Neural Information Processing Systems **8** (1996) 145–151
4. Martoglio, A.M., Miskin, J.W., Smith, S.K., MacKay, D.J.C.: A decomposition model to track gene expression signatures: preview on observer-independent classification of ovarian cancer. Bioinformatics **18**(12) (2002) 1617–1624
5. Griffiths, T., Ghahramani, Z.: Infinite latent feature models and the indian buffet process. Technical Report 1, Gatsby Computational Neuroscience Unit (2005)
6. Ghahramani, Z., Griffiths, T., Sollich, P.: Bayesian nonparametric latent feature models. In: Bayesian Statistics 8, Oxford University Press (2007)
7. Meeds, E., Ghahramani, Z., Neal, R., Roweis, S.: Modeling dyadic data with binary latent factors. In: Neural Information Processing Systems. Volume 19. (2006)
8. Amari, S., Cichocki, A., Yang, H.H.: A new learning algorithm for blind signal separation. In: Advances in Neural Information Processing Systems, 8:757–763. The MIT Press (1996)