Computer Vision and Image Understanding 151 (2016) 90–100

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Non-parametric Bayesian models of response function in dynamic image sequences

Ondřej Tichý*, Václav Šmídl

Institute of Information Theory and Automation, Pod Vodárenskou Věží 4, Prague 8 18208, Czech Republic

ARTICLE INFO

Article history:
Received 11 December 2014
Accepted 21 November 2015

Keywords:
Response function
Blind source separation
Dynamic medical imaging
Probabilistic models
Bayesian methods

ABSTRACT

Estimation of response functions is an important task in dynamic medical imaging. This task arises for example in dynamic renal scintigraphy, where impulse response or retention functions are estimated, or in functional magnetic resonance imaging where hemodynamic response functions are required. These functions can not be observed directly and their estimation is complicated because the recorded images are subject to superposition of underlying signals. Therefore, the response functions are estimated via blind source separation and deconvolution. Performance of this algorithm heavily depends on the used models of the response functions. Response functions in real image sequences are rather complicated and finding a suitable parametric form is problematic. In this paper, we study estimation of the response functions using non-parametric Bayesian priors. These priors were designed to favor desirable properties of the functions, such as sparsity or smoothness. These assumptions are used within hierarchical priors of the blind source separation and deconvolution algorithm. Comparison of the resulting algorithms with these priors is performed on synthetic datasets as well as on real datasets from dynamic renal scintigraphy. It is shown that flexible non-parametric priors improve estimation of response functions in both cases. MATLAB implementation of the resulting algorithms is freely available for download.

1. Introduction

Computer analysis of dynamic image sequences offers an opportunity to obtain information about organ function without invasive intervention. A typical example is replacement of invasive blood sampling by computer analysis of dynamic images [1]. The unknown input function can be obtained by deconvolution of the organ time activity curve and organ response function. Typically, both the input function and the response functions are unknown. Moreover, the time-activity curves are also not directly observed since the recorded images are observed as superposition of multiple signals. The superposition arises e.g. from partial volume effect in dynamic positron emission tomography [2] or dynamic and functional magnetic resonance imaging [3] or from projection of the volume into planar dynamic scintigraphy [4]. Analysis of the dynamic image sequences thus requires to separate the original sources (source images, mean images of active components) and their weights over the time forming the time-activity curves (TACs). The TACs are then decomposed into input function and response functions. Success of the procedure is dependent on the model of the image sequence.

The common model for dynamic image sequences is the factor analysis model [5], which assumes linear combination of the source images and TACs. Another common model is that TAC arise as a convolution of common input function and source specific kernel [6,7]. The common input function is typically the original signal from the blood and the role of convolution kernels vary from application area: impulse response or retention function in dynamic renal scintigraphy [8] or hemodynamic response function in functional magnetic resonance imaging [9]. In this paper, we will refer to the source kernels as the response functions, however other interpretations are also possible.

Analysis of the dynamic image sequences can be done with supervision of experienced physician or technician, who follows recommended guidelines and uses medical knowledge. However, we aim at fully automated approach where the analysis fully depends on the used model. The most sensitive parameter of the analysis is the model of the response functions (i.e. the convolution kernels). Many parametric models of response functions have been proposed, including the exponential model [10] and the piece-wise linear model [11,12]. An obvious disadvantage of the approach is that the real response function may differ from the assumed parametric models. Therefore, more flexible class of models based on non-parametric ideas were proposed such as averaging over region [13], temporal regularization using finite impulse response filters [14], or

* Corresponding author.
E-mail address: otichy@utia.cas.cz (O. Tichý).

http://dx.doi.org/10.1016/j.cviu.2015.11.010
1077-3142/

free-form response functions using automatic relevance determination principle [15].

In this paper, we will study the probabilistic models of response functions using Bayesian methodology within the general blind source separation model [16]. The Bayesian approach was chosen for its inference flexibility and for its ability to incorporate prior information of models [17,18]. We will formulate the prior model for general blind source separation problem with deconvolution [15] where the hierarchical structure of the model allow us to study various versions of prior models of response functions. Specifically, we design different prior models of the response functions with more parameters then the number of points in the unknown response function. The challenge is to regularize the estimation procedure such that all parameters are estimated from the observed data. We will use the approximate Bayesian approach known as the Variational Bayes method [19]. The resulting algorithms are tested on synthetic as well as on real datasets and comparisons with parametric methods are provided.

2. Probabilistic blind source separation with deconvolution

In this Section, we introduce a model of dynamic image sequences. Estimation of the model parameters yields an algorithm for Blind Source Separation and Deconvolution. Prior models of all parameters except for the response functions are described here while the priors for the response functions will be studied in details in the next section.

2.1. Model of observation

Each recorded image is stored as a column vector $\mathbf{d}_j \in \mathbf{R}^{p\times 1}$, $j = 1,\dots,n$, where n is the total number of recorded images. Each vector $\mathbf{d}_j$ is supposed to be an observation of a superposition of r source images $\mathbf{a}_k \in \mathbf{R}^{p\times 1}$, $k = 1,\dots,r$, stored again columnwise. The source images are weighted by their specific activities in time j denoted as $x_{1,j},\dots,x_{r,j} \equiv \bar{\mathbf{x}}_j \in \mathbf{R}^{1\times r}$. Formally,

$$\mathbf{d}_j = \mathbf{a}_1 x_{1,j} + \mathbf{a}_2 x_{2,j} + \cdots + \mathbf{a}_r x_{r,j} + \mathbf{e}_j = A\bar{\mathbf{x}}_j^T + \mathbf{e}_j, \tag{1}$$

where $\mathbf{e}_j$ is the noise of the observation, $A \in \mathbf{R}^{p\times r}$ is the matrix composed from source images as its columns $A = [\mathbf{a}_1,\dots,\mathbf{a}_r]$, and symbol $()^T$ denotes transposition of a vector or a matrix in the whole paper. The Eq. (1) can be rewritten in the matrix form. Suppose the observation matrix $D = [\mathbf{d}_1,\dots,\mathbf{d}_n] \in \mathbf{R}^{p\times n}$ and the matrix with TACs in its columns, $X = [\bar{\mathbf{x}}_1^T,\dots,\bar{\mathbf{x}}_n^T]^T \in \mathbf{R}^{n\times r}$. Note that we will use the bar symbol, $\bar{\mathbf{x}}_k$, to distinguish the kth row of matrix X, while $\mathbf{x}_k$ will be used to denote the $k$the column. Then, the Eq. (1) can be rewritten into the matrix form as

$$D = AX^T + E. \tag{2}$$

The tracer dynamics in each source is commonly described as convolution of common input function, vector $\mathbf{b} \in \mathbf{R}^{n\times 1}$, and source specific response function (convolution kernel, mathematically), vector $\mathbf{u}_k \in \mathbf{R}^{n\times 1}$, $k = 1,\dots,r$ [10,11,20]. Using convolution assumption, each weight $\mathbf{x}_k$ can be rewritten as

$$\mathbf{x}_k = B\mathbf{u}_k, \quad \forall k = 1,\dots,r, \tag{3}$$

where the matrix $B \in \mathbf{R}^{n\times n}$ is composed from elements of input function $\mathbf{b}$ as

$$B = \begin{pmatrix} b_1 & 0 & 0 & 0 \\ b_2 & b_1 & 0 & 0 \\ \dots & b_2 & b_1 & 0 \\ b_n & \dots & b_2 & b_1 \end{pmatrix}. \tag{4}$$

Suppose the aggregation of response functions $U = [\mathbf{u}_1,\dots,\mathbf{u}_r] \in \mathbf{R}^{n\times r}$. Then, $X = BU$ and the model (2) can be rewritten as

$$D = AU^T B^T + E. \tag{5}$$

The task of subsequent analysis is to estimate the matrices A and U and the vector $\mathbf{b}$ from the data matrix D.

2.1.1. Noise model

We assume that the noise has homogeneous Gaussian distribution with zero mean and unknown precision parameter ω, $e_{i,j} = \mathcal{N}(0,\omega^{-1})$. Then, the data model (2) can be rewritten as

$$f(D|A,X,\omega) = \prod_{j=1}^{n} \mathcal{N}(A\bar{\mathbf{x}}_j, \omega^{-1} I_p), \tag{6}$$

where symbol $\mathcal{N}$ denotes Gaussian distribution and I_p is identity matrix of the size given in its subscript. Since all unknown parameters must have their prior distribution in the Variational Bayes methodology, the precision parameter ω has a conjugate prior in the form of the Gamma distribution

$$f(\omega) = \mathcal{G}(\vartheta_0, \rho_0), \tag{7}$$

with chosen constants ϑ_0, ρ_0.

2.2. Probabilistic model of source images

The only assumption on source images is that they are sparse, i.e. only some pixels of source images are non-zeros. The sparsity is achieved using prior model that favors sparse solution depending on data [21]. We will employ the automatic relevance determination (ARD) principle [22] based on joint estimation of the parameter of interest together with its unknown precision. Specifically, each pixel $a_{i,k}$ of each source image has Gaussian prior truncated to positive values (see A.1, denoted as $t\mathcal{N}$ in this paper) with unknown precision parameter $\xi_{i,k}$ which is supposed to have conjugate Gamma prior as

$$f(a_{i,k}|\xi_{i,k}) = t\mathcal{N}(0, \xi_{i,k}^{-1}), \tag{8}$$

$$f(\xi_{i,k}) = \mathcal{G}(\phi_0, \psi_0), \tag{9}$$

for $\forall i = 1,\dots,p, \forall k = 1,\dots,r$, and ϕ_0, ψ_0 are chosen constants. The precisions $\xi_{i,k}$ form the matrix Ξ of the same size as A.

2.3. Probabilistic model of input function

The input function $\mathbf{b}$ is assumed to be a positive vector; hence, it will be modeled as truncated Gaussian distribution to positive values with scaling parameter $\varsigma \in \mathbf{R}$ as

$$f(\mathbf{b}|\varsigma) = t\mathcal{N}(\mathbf{0}_{n,1}, \varsigma^{-1} I_n), \tag{10}$$

$$f(\varsigma) = \mathcal{G}(\zeta_0, \eta_0), \tag{11}$$

where $\mathbf{0}_{n,1}$ denotes zeros matrix of the given size and ζ_0, η_0 are chosen constants.

2.4. Models of response functions

So far, we have formulated the prior models for source images A and input function $\mathbf{b}$ from decomposition of the matrix D. The task of this paper is to propose and study prior models for response functions U as illustrated in Fig. 1. Different choices of the priors on the response functions have strong influence on the results of the analysis which will be studied in the next section.

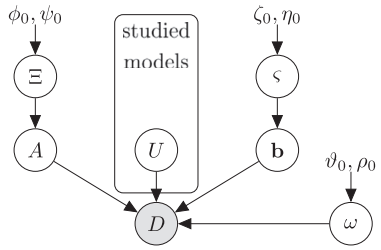

Fig. 1. Hierarchical model for blind source separation with deconvolution problem.

3. Non-parametric prior models of response function

Here, we will formulate several prior models of response functions. Our purpose is not to impose any parametric form as it was done, e.g., in [10,11] but model response function as a free-form curve with only influence from their prior models. The motivation is demonstrated in Fig. 2, where a common parametric model [11] is compared to an example of response function obtained from real data. While the basic form of the response function is correct, exact parametric form of the function would be very complex. Therefore, we prefer to estimate each point on the response function individually. However, this leads to over-parameterization and poor estimates would result without regularization. All models in this Section introduce regularization of the non-parametric response function via modeling of the covariance matrix of the response function.

3.1. Orthogonal prior

The first prior model assumes that each response function $\mathbf{u}_k, k = 1, \ldots, r$, is positive and each response function is weighted by its own precision relevance parameter $\upsilon_k \in \mathbf{R}$ which has a conjugate Gamma prior:

$$f(\mathbf{u}_k|\upsilon_k) = t\mathcal{N}\left(\mathbf{0}_{n,1}, \upsilon_k^{-1} I_n\right), \tag{12}$$

$$f(\upsilon_k) = \mathcal{G}(\alpha_0, \beta_0), \tag{13}$$

for $\forall k = 1, \ldots, r$ and where α_0, β_0 are chosen constants.

The precision parameters υ_k serve for suppression of weak response functions during iterative computation and therefore as parameters responsible for estimation of number of relevant sources.

3.2. Sparse prior

The model with sparse response functions has been introduced in [15]. The key assumption of this model is that the response functions are most likely sparse which is modeled similarly as in case of source images, Section 2.2, using the ARD principle. Here, each element of response function $u_{k,j}$ has its relevance parameter $\upsilon_{k,j}$ which is supposed to be conjugate Gamma distributed. In vector notation, each response function $\mathbf{u}_k$ has its precision matrix $\Upsilon_k = \begin{pmatrix} \upsilon_{k,1} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \upsilon_{k,n} \end{pmatrix}$ with precision parameters $\upsilon_{k,j}$ on its diagonal and zeros otherwise. Then

$$f(\mathbf{u}_k|\Upsilon_k) = t\mathcal{N}\left(\mathbf{0}_{n,1}, \Upsilon_k^{-1}\right), \tag{14}$$

$$f(\upsilon_{k,j}) = \mathcal{G}(\alpha_0, \beta_0), \quad \forall j = 1, \ldots, n, \tag{15}$$

where α_0, β_0 are chosen constants.

This model is extended version of the model (12)–(13) where only one common precision parameter is assumed for each response function. Here, each element of each response function has its own precision parameter which is called ARD principle. The employed ARD principle should suppress the noisy parts of response functions which should lead to clearer response functions and subsequently to clearer TACs.

3.3. Sparse differences prior

Modeling of only sparsity in response functions could possibly lead to arbitrary solution such as very non-smooth curve. The model of differences in response functions allow us to formulate the model favoring smooth response functions which is biologically reasonable requirement. Let us suppose the model of differences of response function $\mathbf{u}_k$, $\nabla\mathbf{u}_k$, where the difference matrix ∇ is defined as

$$\nabla = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & \ddots & 0 \\ 0 & 0 & \ddots & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{16}$$

with ARD prior on each difference using precision parameter $\upsilon_{k,j}$ forming again precision matrix Υ_k; however, with precisions of differences on its diagonal. Formally, the model of differences is

$$f(\nabla\mathbf{u}_k|\Upsilon_k) = \mathcal{N}\left(\mathbf{0}_{n,1}, \Upsilon_k^{-1}\right). \tag{17}$$

We can reformulate the Eq. (17) directly for the response function $\mathbf{u}_k$ as

$$f(\mathbf{u}_k|\Upsilon_k) = \mathcal{N}\left(\nabla\mathbf{0}_{n,1}, \nabla^{-1}\Upsilon_k^{-1}\nabla^{-T}\right), \tag{18}$$

where symbol $()^{-T}$ denotes transpose and inversion of matrix. In model of differences of response function (17), no positivity of response functions could be ensured. However, the direct model of response function (18) has potential benefit since we can a priori assume that the elements of $\mathbf{u}_k$ are positive using truncated normal distribution as

$$f(\mathbf{u}_k|\Upsilon_k) = t\mathcal{N}\left(\mathbf{0}_{n,1}, \nabla^{-1}\Upsilon_k^{-1}\nabla^{-T}\right). \tag{19}$$

The prior model (19) is accompanied by prior model for precisions in the same way as in (15):

$$f(u_{k,j}) = \mathcal{G}(\alpha_0, \beta_0), \quad \forall j = 1, \ldots, n, \tag{20}$$

where α_0, β_0 are chosen constants.

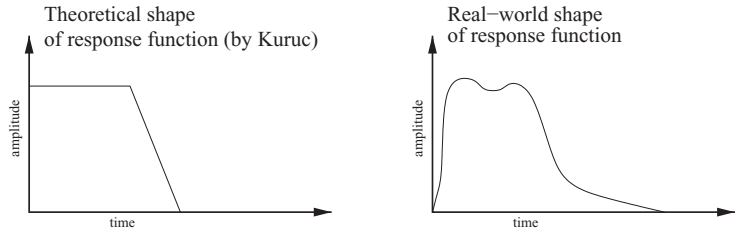

Fig. 2. Example of theoretical shape of response function (by [11]), left, and corresponding real-world shape of convolution kernels, right.

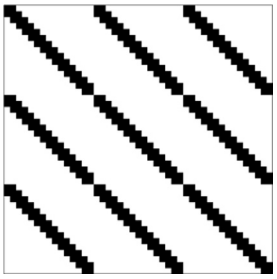

Fig. 3. The used localization matrix L for the first two diagonals. The black pixels denote ones and the white pixels denote zeros. This example is given for $n = 15$ and $r = 3$.

3.4. Wishart prior

So far, we have modeled only the first or the second diagonal of the precision matrix Υ_k. Each of these approaches have its advantages which we would like to generalize into estimation of several diagonals of the prior covariance matrix. However, this is difficult to solve analytically. Instead, we note that it is possible to create the model for the full prior covariance matrix of the response functions as well as their mutual interactions. For this task, we use vectorized form of response functions denoted as $\mathbf{u} \in \mathbf{R}^{nr\times 1}$, $\mathbf{u} = \mathrm{vec}(U) = \left[\mathbf{u}_1^T, \ldots, \mathbf{u}_r^T\right]^T$. This rearranging allow us to model mutual correlation between response functions. The full covariance matrix $\Upsilon \in \mathbf{R}^{nr\times nr}$ can be modeled using Wishart distribution, see A.3, as

$$f(\mathbf{u}|\Upsilon) = t\mathcal{N}\left(\mathbf{0}_{nr,1}, \Upsilon^{-1}\right), \tag{21}$$

$$f(\Upsilon) = \mathcal{W}(\alpha_0 I_{nr}, \beta_0), \tag{22}$$

with scalar prior parameters α_0, β_0.

The advantage of this parametrization is obvious, the full covariance matrix is estimated. The disadvantage is this model is that for estimation nr parameters in vector $\mathbf{u}$, we need to estimate n^2r^2 additional parameters in covariance structure. The problem is regularized by the prior on Y, (22), which is relatively weak regularization with potential side effects. We try to suppress these side effects in the next section.

3.5. Wishart prior with localization

Since restriction of the covariance structure to several diagonals is infeasible in the considered dimensions, we apply an alternative approach known as localization. This techniques originates in data assimilation of atmospheric models [23]. The basic idea of the method is that the most information is localized on the first two diagonals of the matrix Y and its sub-matrices. Hence, we can use Hadamard product, known also as element-wise product, of the original estimates Y and localization matrix L of the same size as the matrix Y. The localization matrix used in this paper for the first two diagonals is illustrated in Fig. 3.

After localization, the model of response functions is the same as in Section 3.4, (21)–(22), however, the estimate of Y, $\widehat{\Upsilon}$, is replaced by

$$\widehat{\Upsilon}_{new} = \widehat{\Upsilon} \circ L, \tag{23}$$

where symbol $\circ$ denotes the Hadamard product. We will show that this localization is a soft version of smoothing of the Wishart model from Section 3.4; however, not so strict as modeling of differences in Section 3.3.

Theoretically, we could employ any conceivable localization as well as smoother version of localization using smooth transitions between ones and zeros; however, this is out of scope of this paper.

Algorithm 1 Bayesian blind source separation algorithm with deconvolution using selected models.

1. Initialization:
   (a) Set prior parameters $\alpha_0, \beta_0, \vartheta_0, \rho_0, \phi_0, \psi_0, \zeta_0, \eta_0$ to non-informative values 10^{-10} or 10^{+10}.
   (b) Set initial values for $\widehat{\Xi}, \widehat{\mathbf{u}}, \widehat{\mathbf{u}^T\mathbf{u}}, \widehat{\Upsilon}, \widehat{\mathbf{b}}, \widehat{\mathbf{b}^T\mathbf{b}}, \widehat{\varsigma}, \widehat{\omega}$.
   (c) Set the initial number of sources r_{max}.
2. Iterate until convergence is reached using computation of shapingparameters and moments from Appendix B:
   (a) Shaping parameters of source images $\mu_{\bar{\mathbf{a}}_i}, \Sigma_{\bar{\mathbf{a}}_i}$and their variances $\psi_i, \phi_i\ \forall i$ using (B.1)–(B.4)and moments $\widehat{A}, \widehat{A^TA}$ using **??**.
   (b) Response functions $\mu_{\mathbf{u}}$ using Eq. (B.10)and $\Sigma_{\mathbf{u}}$ using
   $$\Sigma_{\mathbf{u}} = \left(\widehat{A^TA} \otimes \widehat{\omega}\widehat{B^TB} + R_{\mathbf{u}}\right)^{-1}, \tag{28}$$
   where regularization matrix $R_{\mathbf{u}}$ and hyperparametersaggregated in the estimate $\widehat{\Upsilon}$ depends on version ofthe prior:
      i. Orthogonal: $R_{\mathbf{u}} = I_n \otimes \widehat{\Upsilon}$, $\widehat{\Upsilon}_{k,k} = \frac{\alpha_0+\frac{n}{2}}{\beta_0+\frac{1}{2}\mathrm{tr}\left(\mathbf{u}_k\mathbf{u}_k^T\right)}, k = 1, \ldots, r$,
      ii. Sparse: $R_{\mathbf{u}} = \mathrm{diag}(\mathrm{vec}(\widehat{\Upsilon}))$, $\widehat{\Upsilon} = \mathrm{diag}\left(\frac{\alpha_0+\frac{1}{2}}{\beta_0+\frac{1}{2}\mathrm{diag}\left(\widehat{\mathbf{u}\mathbf{u}^T}\right)}\right)$,
      iii. Sparse Differences: $R_{\mathbf{u}} = (I_r \otimes \nabla)\widehat{\Upsilon}(I_r \otimes \nabla^T)$,
      $\widehat{\Upsilon} = \mathrm{diag}\left(\frac{\alpha_0+\frac{1}{2}}{\beta_0+\frac{1}{2}\mathrm{diag}\left(\nabla^T\widehat{\mathbf{u}\mathbf{u}^T}\nabla\right)}\right)$,
      iv. Wishart: $R_{\mathbf{u}} = \widehat{\Upsilon}$, $\widehat{\Upsilon} = (\beta_0+1)\left(\widehat{\mathbf{u}\mathbf{u}^T} + (\alpha_0 I_{nr})^{-1}\right)^{-1}$,
      v. Wishart with localization: $R_{\mathbf{u}} = \widehat{\Upsilon} \circ L$,
      $\widehat{\Upsilon} = (\beta_0+1)\left(\widehat{\mathbf{u}\mathbf{u}^T} + (\alpha_0 I_{nr})^{-1}\right)^{-1}$,
   Moments $\widehat{\mathbf{u}}, \widehat{\mathbf{u}\mathbf{u}^T}$ are computed from $\mu_{\mathbf{u}}, \Sigma_{\mathbf{u}}$ using Appendix A.2.
   (c) Shaping parameters of the input function $\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}}$ and its variance ζ, η using (B.5)–(B.7) and moments $\widehat{\mathbf{b}}$ using Appendix A.2.
   (d) Shaping parameters of the variance of noise ϑ, ρ using (B.8)–(B.9).
3. Report estimates of source images $\widehat{A}$, response functions $\widehat{U}$,and input function $\widehat{\mathbf{b}}$.

3.6. Variational Bayes approximate solution

The whole probabilistic model is formed by Eqs. (6)–(11), and selected response functions model from Sections 3.1–3.5. The probabilistic model is solved using Variational Bayes (VB) method [16,19] which seeks posterior in the conditionally independent form. We follow methodology described in [19] and identify the following posterior distributions denoted as $\tilde{f}$:

$$\tilde{f}(\bar{\mathbf{a}}_i|D) = t\mathcal{N}(\mu_{\bar{\mathbf{a}}_i}, \Sigma_{\bar{\mathbf{a}}_i}), \quad \tilde{f}(\bar{\boldsymbol{\xi}}_i|D) = \mathcal{G}(\boldsymbol{\phi}_i, \boldsymbol{\psi}_i), \tag{24}$$

$$\tilde{f}(\mathbf{b}|\varsigma) = t\mathcal{N}(\mu_{\mathbf{b}}, \Sigma_{\mathbf{b}}), \quad \tilde{f}(\varsigma|D) = \mathcal{G}(\zeta, \eta), \tag{25}$$

$$\tilde{f}(\mathbf{u}|D) = t\mathcal{N}(\mu_{\mathbf{u}}, \Sigma_{\mathbf{u}}), \quad \tilde{f}(\Upsilon|D) = \mathcal{G}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \text{ or } \mathcal{W}(\Sigma_\Upsilon, \beta), \tag{26}$$

$$\tilde{f}(\omega|D) = \mathcal{G}(\vartheta, \rho), \tag{27}$$

where shaping moments of these posteriors are given in Appendix B. The shaping parameters together with the moments of the distributions (24)–(27) form a set of implicit equations that is solved iteratively. The final algorithm is summarized in Algorithm 1 with

branches for different prior models. The algorithm is proven to converge to local optima. Thus the choice of initial value is critically important.

3.6.1. Algorithm initialization

Convergence of general matrix factorization was studied in [19], where the initialization was random and the only local optima were observed for different choice of initial values of $\widehat{\omega}$. However, the convolution model represents another degree of freedom that yields possible local extrema. Specifically, the most important initial value is that of the input function $\widehat{\mathbf{b}}$. Random generation of this value yields significantly different results.

However, in the considered application we have an underlying assumption that the input function is unimodal with highest value at the beginning of the sequence. Therefore, we choose its initial value as an exponential $\widehat{b_j} = \exp(-\frac{j}{3})$. The exact value of the decrease is not important, the algorithms yields consistent results even for very high values. The initial shapes of the response functions are chosen as as pulses with different lengths in order to cover various dynamics of the sources.

Similarly to the simple matrix factorization, the initial value of the variance of the noise has influence especially on the ability of the algorithm to suppress irrelevant sources. Since the application domain is renal scintigraphy where the noise level is very high, we can reliably estimate the initial noise level using principal component analysis [19] as $\widehat{\omega} = \text{abs}(\frac{p}{\min(\text{eig}(D^TD))})$ where abs() denotes absolute value and eig() denotes eigenvalue vector of given matrix.

All prior parameters in step 1a) of the Algorithm can be chosen to yield non-informative prior, i.e. to extreme numerical values 10^{-10} or 10^{+10}.

3.6.2. Convergence issues

For the above described initialization, the proposed algorithms converge monotonically to the final solution. The only exception is a situation when the input function has peak at different element than the first one. Then, the inverse of the covariance matrix $\Sigma_{\mathbf{u}}$ is poorly conditioned yielding numerical instability. The most sensitive algorithm to this issue is the algorithm with sparse prior. Stable convergence is restored when the inversion in (B.13) is replaced by pseudoinverse or conjugate gradient approximation. While it is possible to use stopping rule, we run the algorithm for a fixed number of iterations, typically 300. In our experience the results of the algorithm do not change significantly after more iterations.

3.6.3. Estimation of the number of sources

The Variational Bayes solution with orthogonal prior has the ability to suppress redundant sources [22]. This is a consequence of the prior form, known as automatic relevance determination [21]. In practice, the algorithms estimates a preselected number of sources r_{max} and if some sources are redundant, their posterior mean converges to zero.

Other priors have also the tendency to suppress the redundant sources but no so strongly as the orthogonal prior from Section 3.1. Heuristic algorithms were designed to detect if the weak signal is relevant or not [15], however, this topic is out of scope of this paper. We will discuss this property on simulated data.

4. Experiments and discussion

We proposed five models of non-parametric response functions within the model of probabilistic blind source separation in Sections 3.1–3.5. The proposed algorithms are denoted using the name of their prior. The proposed algorithms are tested on simulated phantom studies as well as on clinical data from dynamic renal scintigraphy. The comparison with other algorithms for blind source separation with deconvolution is given.

4.1. Algorithms used in comparison

The proposed Algorithms will be tested with other methods that solve the blind source separation and deconvolution problem. The competing methods with parametric forms of either input function or response function are briefly reviewed now.

4.1.1. Blind compartment model separation

The blind compartment model separation (BCMS) method, [12], adopts convolution assumption (3). In addition, the convolution kernels are encouraged to follow a piecewise linear prior model [11]. The prior is composed of a constant plateau from the beginning, followed by a linear slope down to zero. Sparsity of the tissue images is not incorporated in the BCMS model; however, the number of sources is determined by an unknown scalar variance for each tissue image.

4.1.2. Convex analysis of mixtures - compartment modeling algorithm

Interpretation of the superposed signal as a mixture model is used in the Convex analysis of mixtures - compartment modeling (CAM-CM) method [10]. Here, the signal is also modeled as in (5) while the shapes of response functions are assumed to have exponential shape $\exp\left(-k_{\text{ep},j}t\right)$, where $k_{\text{ep},j}$ is the flux rate in source j and t denotes time index. The algorithm is based on identifying pure-volume pixels of each source and clustering pixel time series around them.

4.1.3. Non-negative matrix factor deconvolution

Non-negative matrix Factor Deconvolution (NMFD) [24] is extended version of the classical NMF algorithm [25] where factorization in the form $D \approx AX^T$ is seek with positivity constraints on both, A and X. However, NMF does not take into account relative positions of each spectrum and subsequently discards temporal information. In NMFD, the factorization is seek in the form $D \approx \sum_{t=0}^{n-1} A_t X^{t\rightarrow}$, where operator $()^{t\rightarrow}$ shifts the columns of its argument by t spots to the right.

4.2. Synthetic datasets

We first study performance of the methods on simulated data since we can compare the estimates with the simulated values.

4.2.1. Toy example

Performance of the proposed models of response functions is first studied on a synthetic dataset generated according to the model (5). The size of each image is 50×50 pixels and the number of simulated time points is $n = 50$. We simulate 3 sources in the dataset which are given in Fig. 4, top row, using their source images and response functions together with generated input function $\mathbf{b}$ (top row, right). We generate homogeneous Gaussian noise with standard deviation 0.3 of the signal strength.

The results of the five proposed models are given in Fig. 4 in the row-wise schema. All algorithms run with the initial number of sources $r_{max} = 4$ in order to study behaviors of the algorithm under condition of overestimated number of sources. It can be seen that all methods estimated the source images correctly except the method with sparse prior where the overlap of the first and the third source is estimated as the forth source. The activities in the redundant forth source obtained by the other methods are negligible. Thus, a heuristic procedure in the sense of [15] could be designed to remove them.

The proposed algorithms provide comparable estimate of the source images, however, the estimates of the response function differ, the fifth to the eighth columns, as well as the estimated input function, the ninth column. Note that only the first prior, orthogonal, was not able to respect the sparse character of the modeled response functions, all other priors were able to do so. The visual results are accompanied by the corresponding mean square errors (MSE) summarized in

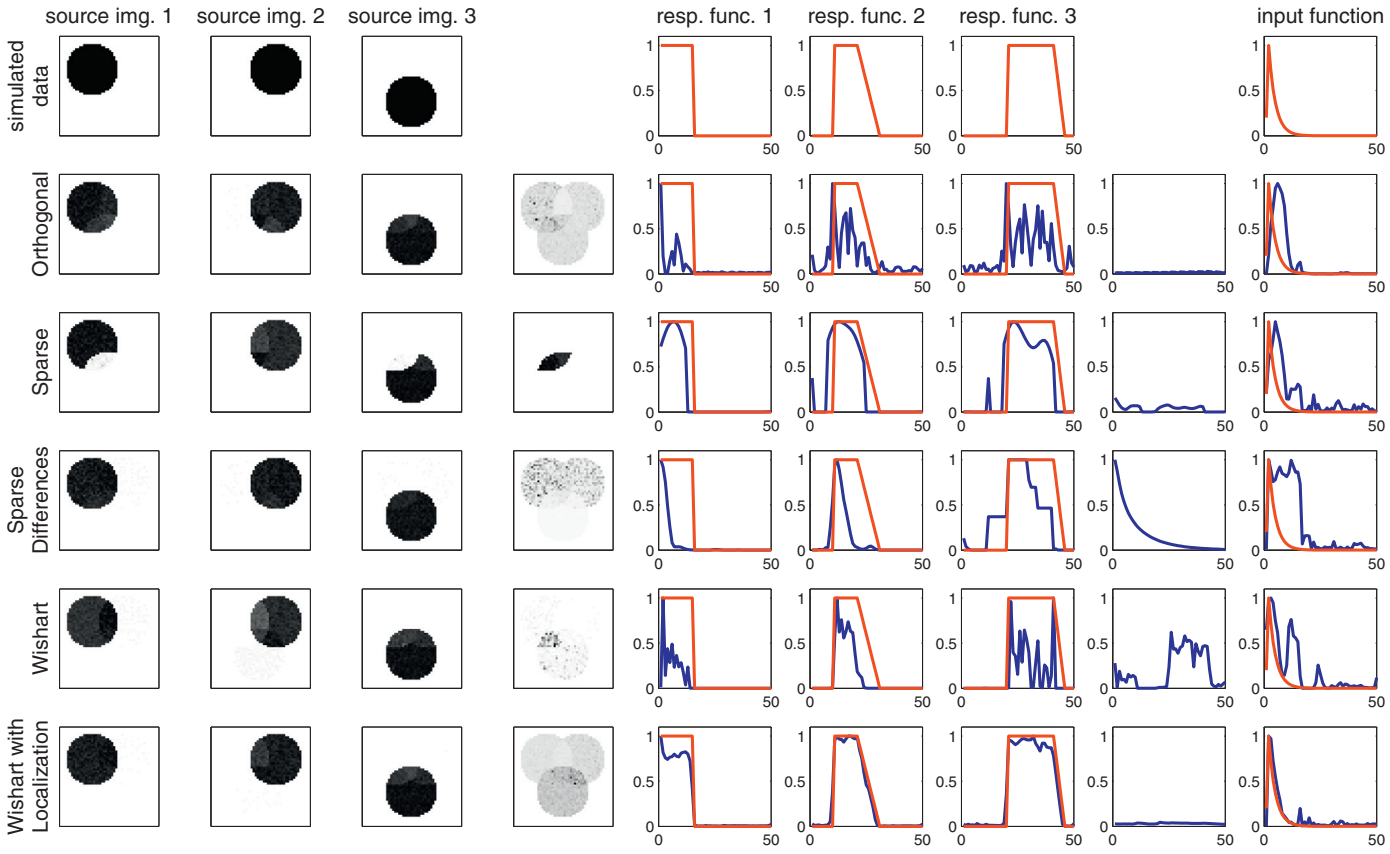

Fig. 4. The results of the five studied methods on synthetic dataset (the first row). The red lines are generated data while the blue lines are estimated results from the respected methods. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 1
Computed mean square errors (MSE) from the simulated data.

Prior model of the response function		Total MSE on U	MSE on **b**
Orthogonal,	Section 3.1	27.34	3.39
Sparse,	Section 3.2	10.47	2.20
Sparse differences,	Section 3.3	23.56	7.26
Wishart,	Section 3.4	26.15	2.71
Wishart with localization,	Section 3.5	1.70	0.28

Table 1 . Here, the MSE is computed between the estimated response functions and their simulated values as well as between the estimated input functions and its simulated value for each method. The Wishart prior with localization outperforms the other ones in estimation of both, response functions and input function.

The effect of localization on the algorithm with the Wishart prior is illustrated in Fig. 5 via sensitivity study of the prior parameter α_0 on the resulting MSE of the response functions. The prior parameter β_0 is selected as 10^{-10} for all cases. For large values of α_0 the results are comparable, however the localized version is improving with decreasing α_0. For values of $\alpha_0 < 1$, the results of the localized version stabilize and become insensitive to the exact value of α_0.

4.2.2. Realistic Monte Carlo phantom from dynamic renal scintigraphy

The algorithms are tested with Monte Carlo simulated anthropomorphic phantom of dynamic renal study [26]. The sequence consists of 120 images. The data are expected to contain three main sources of activity: (i) parenchyma, the outer part of a kidney where the tracer is accumulated at the first, (ii) pelvis, the inner part of a kidney where the accumulation has physiological delay, and (iii) background tissues which is typically active at the beginning of the sequence. However, the data were generated from physiological models which does not generate the curves expected by the simplified linear model. Nevertheless, it provides an additional image sequence with the right kidney without tissue background, attenuation and noise. From this additional sequence, the reference time-activity curve (TAC) of parenchyma can be extracted easily without interference of the contaminating structures. This TAC will be used for comparison of the blind source separation methods.

The results from all five proposed methods as well as those of the competing methods BCMS, CAM-CM, and NMFD algorithms are given in Fig. 6 in row-wise scheme while all algorithms have preselected $r_{max} = 4$. Note that TACs for each method are displayed instead of response functions since ground truth is given as TAC of parenchyma directly. Estimated source images of parenchyma can be found in the first column accompanied by estimated TACs of parenchyma in the fifth column, blue line, while the ground truth is displayed using red line. We conclude that the estimate of the method with Wishart prior with localization fits to the ground truth better than other methods while the worst estimate is provided by NMFD algorithm which is clearly not able to cope with such a noisy data. Estimates of the source displayed in the second column can be interpreted as estimated activity of the pelvis and those in the third column as those of the tissue background. The corresponding TACs are displayed in the same order, Fig. 6. Note that the CAM-CM algorithm was not able to separate these sources while other algorithms were. On this data, the best results are those of the the method with Wishart prior with localization. Moreover, they are closer to physiological expectation then results from other methods since the TAC of the pelvis is realistic (although it can not be validated without ground truth data).

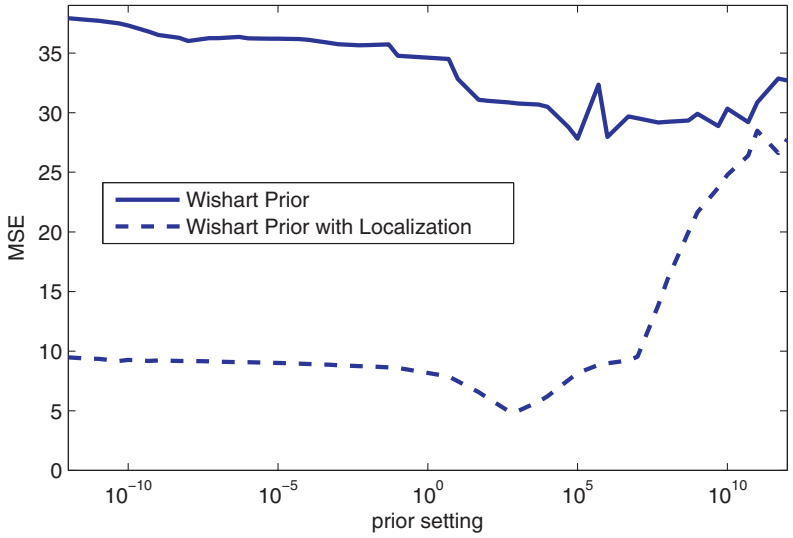

Fig. 5. Sensitivity study of MSE of estimated on the parameter α_0 of the Wishart prior and the Wishart prior with localization.

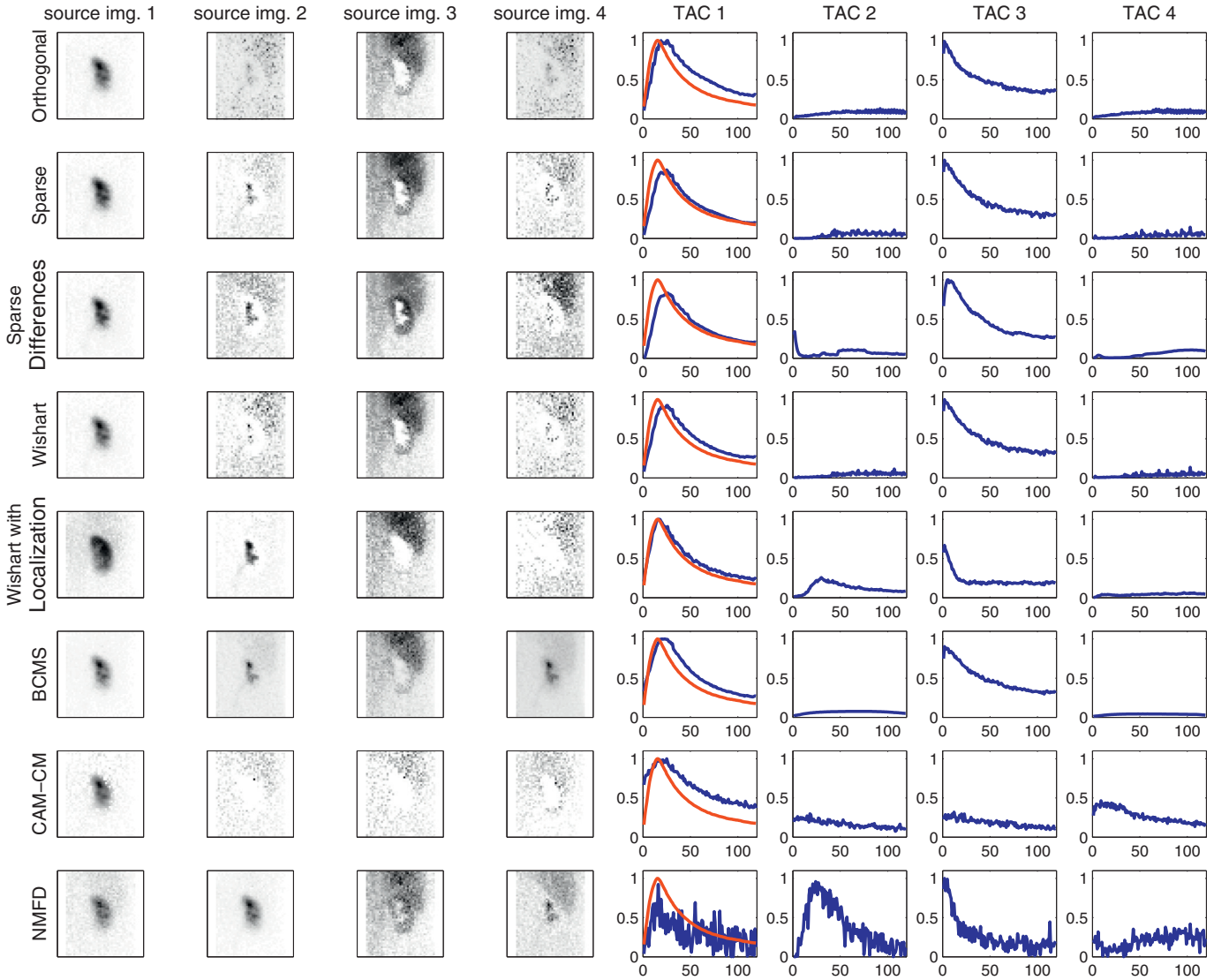

Fig. 6. Estimated source images (columns 1–4) and time-activity curves (columns 5–8) using priors: Orthogonal, Sparse, Sparse differences, Wishart, Wishart with localization and state-of-the-art algorithms: BCMS, CAM-CM, and NMFD.

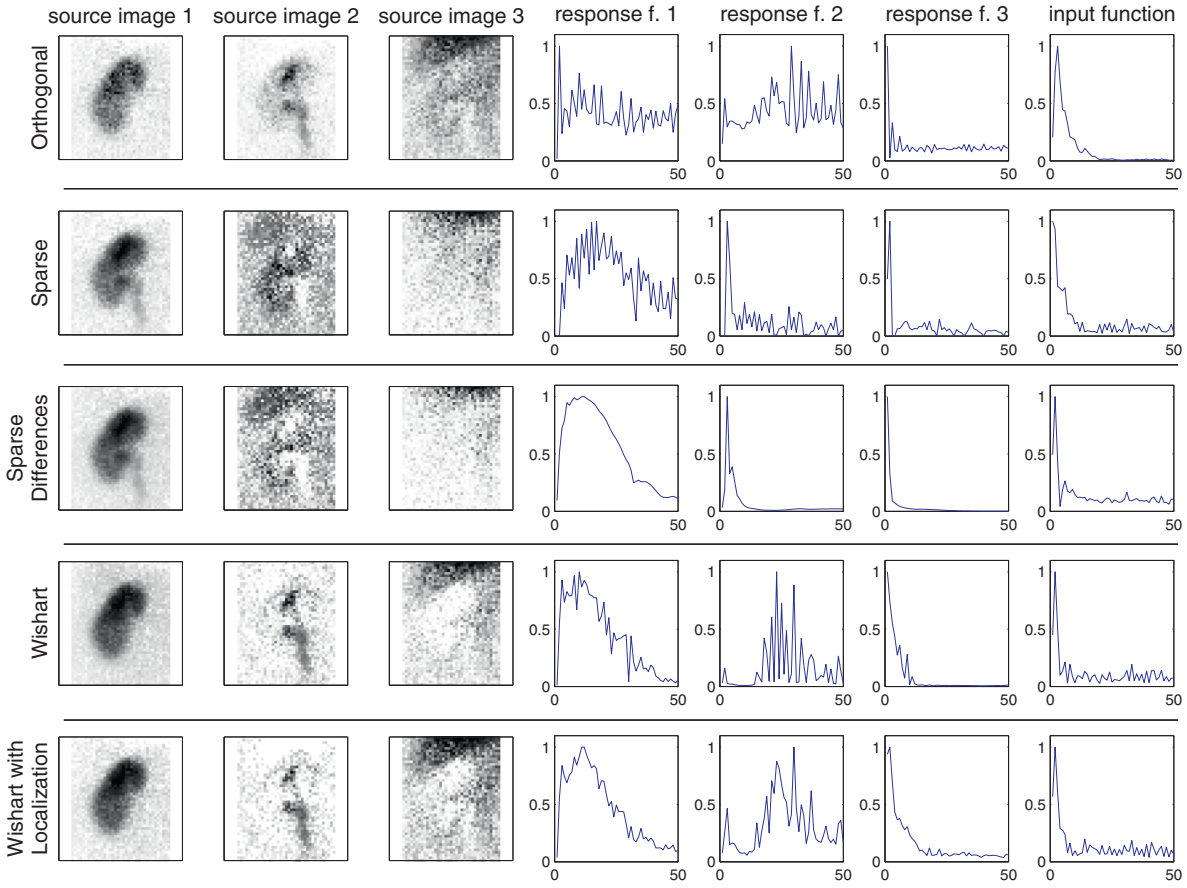

Fig. 7. Estimated source images (columns 1–3), response functions (columns 4–6), and input functions (column 7) using priors: Orthogonal, Sparse, Sparse differences, Wishart, Wishart with localization.

4.3. Dynamic renal scintigraphy dataset

A database with clinical data is publicly available[1] consisting of 107 sequences from dynamic renal scintigraphy. Each sequence has spatial resolution 128 × 128 pixels and the number of images is 180 taken with sampling period 10 s. For detailed clinical description see [27]. These data are assumed to be suitable for analysis by the proposed model. The only difference in model of the noise which is Poisson distributed. The assumption of homogeneous Gaussian noise (6) can be achieved by asymptotic scaling known as the correspondence analysis [28] which transforms the original data D_{orig} as

$$d_{ij} = \frac{d_{ij,orig}}{\sqrt{\sum_{i=1}^{p} d_{ij,orig} \sum_{j=1}^{n} d_{ij,orig}}}. \tag{29}$$

The proposed methods are applied to the scaled data.

Validation of the methods on this dataset is very challenging since the only provided results of analysis is the relative renal function coefficient which is estimated manually by an experienced physician. Since we do not have ground truth data for the source images and time-activity-curves, we first discuss qualitative differences of the separation on a selected sequence.

4.3.1. Example dataset from dynamic renal scintigraphy

In general, the proposed methods may provide very similar results especially when the data are informative, i.e. the organs have strong activity with only minor overlap. The most visible differences between the methods are on data with very poor signal-to-noise ratio. These data are obtained when the kidney is malfunctioning and are thus most important from clinical point of view. As an example, we show results of analysis of 50 frames of dataset number 42 where different methods yield distinct results. The results are shown in Fig. 7 using the estimated source images (columns 1–3), the estimated related response functions (columns 4–6), and the estimated input function (column 7). Note that the sparse and the sparse differences priors were not able to separate the pelvis which is mixed with the parenchyma in the first column while the orthogonal prior estimated the source images reasonably; however, the response functions of the parenchyma and the pelvis are clearly mixed. The Wishart-based priors, Wishart and Wishart with localization, were able to separate the parenchyma and the pelvis correctly together with meaningful estimates of their response functions. The main difference between the Wishart and the Wishart with localization priors is in smoothness. The estimated response functions from the Wishart prior with localization better matches the physiological expectations than the estimates from the Wishart model. In this case, the use of more complex prior models significantly outperform the simpler models.

Naturally it is possible to find sequences where other methods seems to correspond better with biological expectations, hence it is not possible to conclude that one prior is systematically better than the other.

4.3.2. Differential renal function estimation

The database contains the result of manual analysis in the form of diagnostic parameter called differential renal function (DRF) provided by an experienced physician using standard clinical procedures. Basically, the DRF is a percentual performance of each kidney, hence, it is a relative number. It is a subjective value since the

[1] Database of dynamic renal scintigraphy, http://www.dynamicrenalstudy.org (accessed: 1st December 2014).

Table 2
The number of results which differ from physician's results less than 5%, less than 10%, and more or equal than 10% computed for all 99 patients.

Algorithm		< 5%	< 10%	≧ 10%
Orthogonal,	Section 3.1	68	88	11
Sparse,	Section 3.2	70	90	9
Sparse differences,	Section 3.3	71	89	10
Wishart,	Section 3.4	62	87	12
Wishart with localization,	Section 3.5	71	88	11
BCMS		59	82	17
CAM-CM		48	63	36
NMFD		46	76	23

resulting DRF may differ from physician to physician depending on the used technique and analyzing procedures; however, it allows us to quantitatively compare performance of all methods based on proximity of the estimated DRF to the values from the physician.

The same ROIs as in Section 4.3.1 were placed on the left and on the right kidney and DRF of the left kidney were computed according to $\text{DRF}_L = \frac{P_l}{P_l+P_r} \times 100\%$, where P_l is the total activity of the left parenchyma, and P_r is the total activity of the right parenchyma. Therefore, only sequences with both kidneys were taken into account, i.e. 99 sequences. Historically, the DRF is computed only from the initial phase of sequence where only parenchyma is activated, this phase is selected manually in our experiment. The number of sources is set to 2 for all algorithms since we expect separation of only the parenchyma and the background.

For each sequence we compute the DRF using all competing methods and compare it with the reference value from the database denoted as DRF_L^{gt}. The results are summarized in Table 2 using the number of tested sequences for which the difference $|\text{DRF}_L - \text{DRF}_L^{gt}|$ is lower than 5%, lower than 10%, and more or equal to 10%.

The results suggest that the methods with non-parametric model of response functions have significantly better performance than those with parametric model on this specific task. We conclude that the differences between non-parametric models are under statistical significance.

This improvement was achieved at the prize of higher computation time. Analysis of one sequence takes approximately 20 seconds on standard PC for the proposed methods. In comparison, analysis of the same sequence by the CAM-CM or NMFD takes only few seconds.

5. Conclusion

A common model in functional analysis of dynamic image sequences assumes that the observed images arise from superposition of the original source images weighted by their time-activity curves. Each time-activity curve is assumed to be a result of common input function and source-specific response function, both unknown. Estimation of the model parameters yields an algorithm for blind source separation and deconvolution. The focus of this study is the prior model of the response functions while the models of the source images and the input function are the same. We propose five prior models of the response functions. The first three prior models are based on automatic relevance determination principle on the whole response functions, on each element of the response function, and on the differences between elements of the response functions, respectively. The forth model is based on full model of covariance matrix using Wishart distribution while the fifth model is based on the same prior; however, with additional localization within the deconvolution algorithm. The advantage of all five models is their flexibility in estimation of various shapes of response functions since we do not impose any parametric form of them. The formulated probabilistic models in the form of hierarchical priors are solved using the Variational Bayes methodology.

The performance of the proposed methods is tested on simulated datasets as well as on real datasets from dynamic renal scintigraphy. On the simulated data, the model with Wishart prior with localization provides the best results in the sense of mean square error from the simulated values. However, evaluation on real clinical data is more complex since no ground truth values are available. The method with Wishart prior with localization have subjectively better results on sequences with weak signal-to-noise ratio. However, on large comparison with 99 studies the results of all proposed methods with no-parametric priors are comparable. In this study, all proposed non-parametric methods provide better results than any other parametric methods. Notably, the proposed methods have no domain-specific assumptions; hence, they can be used in other tasks in dynamic medical imaging. The MATLAB implementation of all methods is available for download in http://www.utia.cz/bss_rf_priors/.

Acknowledgment

This work was supported by the Czech Science Foundation, grant no. GA13-29225S.

Appendix A. Required probability distributions

A.1. Truncated normal distribution

Truncated normal distribution, denoted as $t\mathcal{N}$, of a scalar variable x on interval $[a; b]$ is defined as

$$t\mathcal{N}(\mu, \sigma, [a, b]) = \frac{\sqrt{2}\exp((x-\mu)^2)}{\sqrt{\pi\sigma}(erf(\beta) - erf(\alpha))}\chi_{[a,b]}(x), \tag{A.1}$$

where $\alpha = \frac{a-\mu}{\sqrt{2\sigma}}$, $\beta = \frac{b-\mu}{\sqrt{2\sigma}}$, function $\chi_{[a,b]}(x)$ is a characteristic function of interval $[a, b]$ defined as $\chi_{[a,b]}(x) = 1$ if $x \in [a, b]$ and $\chi_{[a,b]}(x) = 0$ otherwise. erf() is the error function defined as $\text{erf}(t) = \frac{2}{\sqrt{\pi}}\int_0^t e^{-u^2}\text{d}u$.

The moments of truncated normal distribution are

$$\widehat{x} = \mu - \sqrt{\sigma}\frac{\sqrt{2}[\exp(-\beta^2) - \exp(-\alpha^2)]}{\sqrt{\pi}(\text{erf}(\beta) - \text{erf}(\alpha))}, \tag{A.2}$$

$$\widehat{x^2} = \sigma + \mu\widehat{x} - \sqrt{\sigma}\frac{\sqrt{2}[b\exp(-\beta^2) - a\exp(-\alpha^2)]}{\sqrt{\pi}(\text{erf}(\beta) - \text{erf}(\alpha))}. \tag{A.3}$$

A.2. Multivariate truncated normal distribution

Truncation of the multivariate Normal distribution of the vector $\mathbf{x}$, $\mathbf{x} \sim t\mathcal{N}(\boldsymbol{\mu}, \Sigma_{\mathbf{x}}, [a, b])$, is formally simple, however, analytically intractable. Hence, we approximate the moments of the vector $\mathbf{x}$of the truncated Normal distribution using moments of

$$\tilde{\mathbf{x}} \sim t\mathcal{N}(\boldsymbol{\mu}, \text{diag}(\boldsymbol{\sigma}_{\mathbf{x}}), [a, b]), \tag{A.4}$$

where $\boldsymbol{\sigma}_{\mathbf{x}}$ is a vector of diagonal elements of $\Sigma_{\mathbf{x}}$ corresponding to the approximation of the posterior by a product of marginals (A.1) with mean value $\widehat{\mathbf{x}}$ with elements given by (A.2) and $\widehat{\mathbf{x}\mathbf{x}^T} = \widehat{\mathbf{x}}\widehat{\mathbf{x}}^T + \text{diag}(\hat{\boldsymbol{\sigma}})$, where $\hat{\sigma}_i = \widehat{x_i^2} - \hat{x}_i\hat{x}_i$. However, it may be too coarse approximation since it ignores covariance of the elements. An alternative is to approximate

$$\widehat{\mathbf{x}\mathbf{x}^T} = \widehat{\mathbf{x}}\widehat{\mathbf{x}}^T + \text{diag}(\mathbf{o})\Sigma_{\mathbf{x}}\text{diag}(\mathbf{o}), \tag{A.5}$$

where $\mathbf{o}$ is a vector of elements $o_i = \hat{\sigma}_i^{1/2}\sigma_i^{-1/2}$. Heuristics (A.5) is motivated by the observation that for a Normal distribution with the main mass far from the truncation lines, $o_i \to 1$ and (A.5) becomes equivalent to the moment of the non-truncated Normal distribution.

A.3. Wishart distribution

Wishart distribution $\mathcal{W}$ of the positive-definite matrix $X \in \mathbf{R}^{p\times p}$ is defined as

$$\mathcal{W}_p(\Sigma,\nu) = |X|^{\frac{\nu-p-1}{2}} 2^{-\frac{\nu p}{2}} |\Sigma|^{-\frac{\nu}{2}} \Gamma_p^{-1}\left(\frac{\nu}{2}\right) \exp\left(-\frac{1}{2}\mathrm{tr}\left(\Sigma^{-1}X\right)\right), \tag{A.6}$$

where $\Gamma_p\left(\frac{\nu}{2}\right)$ is the gamma function. The required moment is:

$$\widehat{X} = \nu\Sigma. \tag{A.7}$$

Appendix B. Shaping parameters of posteriors

Shaping parameters of posterior distributions are given as:

$$\Sigma_{\bar{\mathbf{a}}_i} = \left(\widehat{\omega}\sum_{j=1}^{n}(\widehat{\bar{\mathbf{x}}_j^T\bar{\mathbf{x}}_j}) + \mathrm{diag}(\widehat{\boldsymbol{\xi}}_i)\right)^{-1}, \tag{B.1}$$

$$\mu_{\bar{\mathbf{a}}_i} = \Sigma_{\bar{\mathbf{a}}_i}\widehat{\omega}\sum_{j=1}^{n}(\widehat{\bar{\mathbf{x}}}_j d_{i,j}), \tag{B.2}$$

$$\boldsymbol{\phi}_i = \phi_0 + \frac{1}{2}\mathbf{1}_{r,1}, \tag{B.3}$$

$$\boldsymbol{\psi}_i = \psi_0 + \frac{1}{2}\mathrm{diag}\left(\widehat{\bar{\mathbf{a}}_i^T\bar{\mathbf{a}}_i}\right), \tag{B.4}$$

$$\Sigma_{\mathbf{b}} = \left(\widehat{\varsigma} I_n + \widehat{\omega}\sum_{i,j=1}^{r}(\widehat{\bar{\mathbf{a}}_i^T\bar{\mathbf{a}}_j})\left(\sum_{k,l=0}^{n-1}\Delta_k^T\Delta_l\widehat{u_{k+1,j}u_{l+1,i}}\right)\right)^{-1}, \tag{B.5}$$

$$\mu_{\mathbf{b}} = \Sigma_{\mathbf{b}}\widehat{\omega}\sum_{k=1}^{r}\left(\sum_{j=0}^{n-1}\Delta_j\widehat{u_{j+1,k}}\right)^T D^T\widehat{\mathbf{a}}_k, \tag{B.6}$$

$$\zeta = \zeta_0 + \frac{n}{2},\quad \eta = \eta_0 + \frac{1}{2}\mathrm{tr}\left(\widehat{\mathbf{b}^T\mathbf{b}}\right), \tag{B.7}$$

$$\vartheta = \vartheta_0 + \frac{np}{2}, \tag{B.8}$$

$$\rho = \rho_0 + \frac{1}{2}\mathrm{tr}\left(DD^T - \widehat{A}\widehat{X}^TD^T - D\widehat{X}\widehat{A}^T\right) + \frac{1}{2}\mathrm{tr}\left(\widehat{A^TA}\widehat{X^TX}\right). \tag{B.9}$$

Here, $\widehat{x}$ denotes a moment of respective distribution, tr() denotes a trace of argument, diag() denotes a square matrix with argument vector on diagonal and zeros otherwise or a vector composed from diagonal element of argument matrix, $\mathbf{1}_{n,1}$ denotes the matrix of ones of dimension $n \times 1$, the auxiliary matrix $\Delta_k \in \mathbf{R}^{n\times n}$ is defined as $(\Delta_k)_{i,j} = \begin{cases}1, & \text{if } i-j=k,\\ 0, & \text{otherwise,}\end{cases}$, and standard moments of required probability distributions are given A.1 and A.3 and, e.g., in the appendix of [19].

The shaping parameters for response functions are given in following subsections while the parameter $\mu_{\mathbf{u}}$ is common for all methods as

$$\mu_{\mathbf{u}} = \Sigma_{\mathbf{u}}\left(\widehat{A^TA}\otimes\widehat{\omega}\widehat{B^TB}\right)\mathrm{vec}\left(\widehat{B^TB}^{-1}\widehat{B}^TD^T\widehat{A}\widehat{A^TA}^{-1}\right). \tag{B.10}$$

B.1. Shaping parameters for orthogonal prior

$$\Sigma_{\mathbf{u}} = \left(\widehat{A^TA}\otimes\widehat{\omega}\widehat{B^TB} + I_n\otimes\widehat{\Upsilon}\right)^{-1}, \tag{B.11}$$

$$\alpha_k = \alpha_0 + \frac{n}{2},\ \beta_k = \beta_0 + \frac{1}{2}\mathrm{tr}\left(\widehat{\mathbf{u}_k\mathbf{u}_k^T}\right),\ \widehat{\Upsilon} = \mathrm{diag}\left(\frac{\boldsymbol{\alpha}}{\boldsymbol{\beta}}\right), \tag{B.12}$$

B.2. Shaping parameters for sparse prior

$$\Sigma_{\mathbf{u}} = \left(\widehat{A^TA}\otimes\widehat{\omega}\widehat{B^TB} + \widehat{\Upsilon}\right)^{-1}, \tag{B.13}$$

$$\boldsymbol{\alpha} = \alpha_0 + \frac{1}{2}\mathbf{1}_{nr,1},\ \boldsymbol{\beta} = \beta_0 + \frac{1}{2}\mathrm{diag}\left(\widehat{\mathbf{u}\mathbf{u}^T}\right),\ \widehat{\Upsilon} = \mathrm{diag}\left(\frac{\boldsymbol{\alpha}}{\boldsymbol{\beta}}\right), \tag{B.14}$$

B.3. Shaping parameters for sparse differences prior

$$\Sigma_{\mathbf{u}} = \left(\widehat{A^TA}\otimes\widehat{\omega}\widehat{B^TB} + (I_r\otimes\nabla)\widehat{\Upsilon}(I_r\otimes\nabla^T)\right)^{-1}, \tag{B.15}$$

$$\boldsymbol{\alpha} = \alpha_0 + \mathbf{1}_{n,1}\frac{1}{2},\ \boldsymbol{\beta} = \beta_0 + \frac{1}{2}\mathrm{diag}\left(\nabla^T\widehat{\mathbf{u}\mathbf{u}^T}\nabla\right),\ \widehat{\Upsilon} = \mathrm{diag}\left(\frac{\boldsymbol{\alpha}}{\boldsymbol{\beta}}\right), \tag{B.16}$$

B.4. Shaping parameters for Wishart prior

$$\Sigma_{\mathbf{u}} = \left(\widehat{A^TA}\otimes\widehat{\omega}\widehat{B^TB} + \widehat{\Upsilon}\right)^{-1}, \tag{B.17}$$

$$\Sigma_{\Upsilon} = \left(\widehat{\mathbf{u}\mathbf{u}^T} + (\alpha_0 I_{nr})^{-1}\right)^{-1},\ \beta = \beta_0 + 1,\ \widehat{\Upsilon} = \beta\Sigma_{\Upsilon}, \tag{B.18}$$

B.5. Shaping parameters for Wishart prior with localization

$$\Sigma_{\mathbf{u}} = \left(\widehat{A^TA}\otimes\widehat{\omega}\widehat{B^TB} + \widehat{\Upsilon}\circ L\right)^{-1}, \tag{B.19}$$

$$\Sigma_{\Upsilon} = \left(\widehat{\mathbf{u}\mathbf{u}^T} + (\alpha_0 I_{nr})^{-1}\right)^{-1},\ \beta = \beta_0 + 1,\ \widehat{\Upsilon} = \beta\Sigma_{\Upsilon}. \tag{B.20}$$

References

[1] B. Lanz, C. Poitry-Yamate, R. Gruetter, Image-derived input function from the vena cava for 18F-FDG PET studies in rats and mice, J. Nucl. Med. 55 (8) (2014) 1380–1388.

[2] M. Margadán-Méndez, A. Juslin, S. Nesterov, K. Kalliokoski, J. Knuuti, U. Ruotsalainen, ICA based automatic segmentation of dynamic cardiac PET images., Inf. Technol. Biomed. IEEE Trans. 14 (3) (2010) 795–802.

[3] L. Chaari, T. Vincent, F. Forbes, M. Dojat, P. Ciuciu, Fast joint detection-estimation of evoked brain activity in event-related fMRI using a variational approach, Med. Imaging IEEE Trans. 32 (5) (2013) 821–837.

[4] R. Di Paola, J. Bazin, F. Aubry, A. Aurengo, F. Cavailloles, J. Herry, E. Kahn, Handling of dynamic sequences in nuclear medicine, Nucl. Sci. IEEE Trans. 29 (4) (1982) 1310–1321.

[5] A. Martel, A. Moody, S. Allder, G. Delay, P. Morgan, Extracting parametric images from dynamic contrast-enhanced mri studies of the brain using factor analysis, Med. Image Anal. 5 (1) (2001) 29–39.

[6] J. Fleming, P. Kemp, A comparison of deconvolution and the Patlak-Rutland plot in renography analysis, J. Nucl. Med. 40 (9) (1999) 1503.

[7] T. Taxt, R. Jirik, C.B. Rygh, R. Gruner, M. Bartos, E. Andersen, F.-R. Curry, R.K. Reed, Single-channel blind estimation of arterial input function and tissue impulse response in dce-mri, Biomed. Eng. IEEE Trans. 59 (4) (2012) 1012–1021.

[8] E. Durand, M. Blaufox, K. Britton, O. Carlsen, P. Cosgriff, E. Fine, J. Fleming, C. Nimmon, A. Piepsz, A. Prigent, et al., International Scientific Committee of Radionuclides in Nephrourology (ISCORN) consensus on renal transit time measurements, in: Proceedings of the Seminars in nuclear medicine, 38, Elsevier, 2008, pp. 82–102.
[9] M.A. Lindquist, J. Meng Loh, L.Y. Atlas, T.D. Wager, Modeling the hemodynamic response function in fmri: efficiency, bias and mis-modeling, Neuroimage 45 (1) (2009) S187–S198.
[10] L. Chen, P. Choyke, T.-H. Chan, C.-Y. Chi, G. Wang, Y. Wang, Tissue-specific compartmental analysis for dynamic contrast-enhanced MR imaging of complex tumors, Med. Imaging IEEE Trans. 30 (12) (2011) 2044–2058.
[11] A. Kuruc, W. Caldicott, S. Treves, Improved Deconvolution Technique for the Calculation of Renal Retention Functions., Comp. and Biomed. Res. 15 (1) (1982) 46–56.
[12] O. Tichý, V. Šmídl, M. Šámal, Model-based extraction of input and organ functions in dynamic scintigraphic imaging, Comput. Methods Biomech. Biomed. Eng. Imaging Vis. (2014), doi:10.1080/21681163.2014.916229.(in print)
[13] J. Kershaw, S. Abe, K. Kashikura, X. Zhang, I. Kanno, A bayesian approach to estimating the haemodynamic response function in event-related fmri, Neuroimage 11 (5) (2000) S474.
[14] C. Goutte, F.A. Nielsen, L.K. Hansen, Modeling the hemodynamic response in fMRI using smooth fir filters, Med. Imaging IEEE Trans. 19 (12) (2000) 1188–1201.
[15] O. Tichý, V. Šmídl, Bayesian blind separation and deconvolution of dynamic image sequences using sparsity priors, Med. Imaging IEEE Trans. 34 (1) (2015) 258–266.
[16] J. Miskin, Ensemble Learning for Independent Component Analysis, University of Cambridge, 2000 Ph.D. thesis.
[17] M. Woolrich, Bayesian inference in fMRI, NeuroImage 62 (2) (2012) 801–810.
[18] D.M. Steinberg, O. Pizarro, S.B. Williams, Hierarchical Bayesian models for unsupervised scene understanding, Comput. Vis. Image Underst. 131 (2015) 128–144.
[19] V. Šmídl, A. Quinn, The Variational Bayes Method in Signal Processing, Springer, 2006.
[20] B. Diffey, F. Hall, J. Corfield, The 99mTc-DTPA dynamic renal scan with deconvolution analysis, J. Nucl. Med. 17 (5) (1976) 352.
[21] M. Tipping, Sparse Bayesian learning and the relevance vector machine, J. Mach. Learn. Res. 1 (2001) 211–244.
[22] C. Bishop, M. Tipping, Variational relevance vector machines, in: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, 2000, pp. 46–53.
[23] T. Hamill, J. Whitaker, C. Snyder, Distance-dependent filtering of background error covariance estimates in an ensemble Kalman filter, Mon. Weather Rev. 129 (11) (2001) 2776–2790.
[24] P. Smaragdis, Non-negative matrix factor deconvolution; extraction of multiple sound sources from monophonic inputs, in: Proceedings of the Independent Component Analysis and Blind Signal Separation, Springer, 2004, pp. 494–499.
[25] D.D. Lee, H.S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401 (6755) (1999) 788–791.
[26] G. Brolin, K. Gleisner, M. Ljungberg, Dynamic 99mTc-MAG3 renography: images for quality control obtained by combining pharmacokinetic modelling, an anthropomorphic computer phantom and monte carlo simulated scintillation camera imaging, Phys. Med. Biol. 58 (10) (2013) 3145.
[27] M. Šámal, J. Valoušek, Clinically documented data set of dynamic renal scintigraphy for clinical audits and quality assurance of nuclear medicine software, in: Eur. J. N. Med. Mol. Imaging, 39, Springer, 2012, pp. S170–S171.
[28] H. Benali, I. Buvat, F. Frouin, J. Bazin, R. Paola, A statistical model for the determination of the optimal metric in factor analysis of medical image sequences (FAMIS), Phys. Med. Biol. 38 (1993) 1065.