Noname manuscript No.
(will be inserted by the editor)

Unmixing of Hyperspectral Images using Bayesian Nonnegative Matrix Factorization with Volume Prior

Morten Arngren · Mikkel N. Schmidt · Jan Larsen

Received: date / Accepted: date

Abstract Hyperspectral imaging can be used in assessing the quality of foods by decomposing the image into constituents such as protein, starch, and water. Observed data can be considered a mixture of underlying characteristic spectra (endmembers), and estimating the constituents and their abundances requires efficient algorithms for spectral unmixing. We present a Bayesian spectral unmixing algorithm employing a volume constraint and propose an inference procedure based on Gibbs sampling. We evaluate the method on synthetic and real hyperspectral data of wheat kernels. Results show that our method perform as good or better than existing volume constrained methods. Further, our method gives credible intervals for the endmembers and abundances, which allows us to asses the confidence of the results.

Keywords Bayesian source separation, Hyperspectral image analysis, Volume regularization, Gibbs sampling

Morten Arngren
Technical University of Denmark, DTU Informatics, Bldg. 321, Richard Petersens Plads, DK-2800 Lyngby and FOSS Analytical A/S, Slangerupgade 69, DK-3400 Hillerød
E-mail: ma@imm.dtu.dk, moa@foss.dk

Mikkel N. Schmidt
Technical University of Denmark, DTU Informatics, Bldg. 321, Richard Petersens Plads, DK-2800 Lyngby
E-mail: mns@imm.dtu.dk

Jan Larsen
Technical University of Denmark, DTU Informatics, Bldg. 321, Richard Petersens Plads, DK-2800 Lyngby
E-mail: jl@imm.dtu.dk

1 Introduction

Classic image acquisition and analysis is based on three color bands (red, green, and blue) which is sufficient for human visualization. In the context of identifying or extracting material constituents of e.g. foods, these three wide channels are rarely enough. Hyperspectral image analysis can include more than 100 channels and hence provides the opportunity to

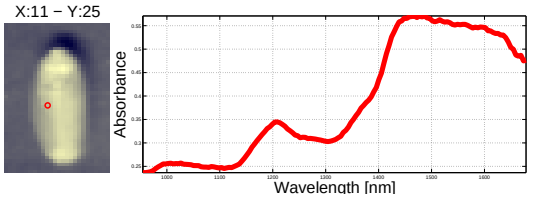


Fig. 1: Pseudo RGB image of a wheat kernel and the associated pre-processed spectrum for the selected pixel.

capture detailed spectral information required in analysis of foods. Figure 1 illustrates a hyperspectral image of a wheat kernel with a corresponding pre-processed spectrum from 950−1650nm.

In a hyperspectral image acquisition system based on light transmission, mixing of the materials constituents can be considered linear according to Lambert-Beer's law. In a reflectance spectroscopy system, which we use in the experiments in this paper, non-linear mixing can occur, when incident light interacts with several constituent materials. For simplicity, however, it is reasonable to assume [6] that the mixing process is predominantly linear and that non-linear effects can be neglected.

Given a set of N preprocessed M-dimensional mixed spectra, stored as columns in a matrix $\boldsymbol{X} \in \mathbb{R}^{M\times N}$, the unmixing problem consists of estimating the non-negative spectral signature of the pure constituent components (endmembers) as well as their relative contributions (fractional abundances) for each of the N spectra. The linear mixing can then be expressed as a rank K linear matrix factorization,

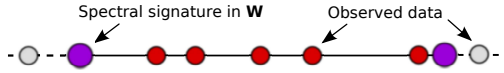
$$\boldsymbol{X} = \boldsymbol{W}\boldsymbol{H} + \boldsymbol{\epsilon}, \tag{1}$$

where K is the number of endmembers. The non-negative spectral signatures of these endmembers are contained in the columns of $\boldsymbol{W} \in \mathbb{R}_+^{M\times K}$ and $\boldsymbol{H} \in \mathbb{R}_+^{K\times N}$ holds the fractional abundances for the N elements. The matrix $\boldsymbol{\epsilon}$ denotes the residual noise. Each of the observed pixels can thus be considered a mix of latent pure constituents. In foods, these constituents are typically water, protein, starch, oil, etc.

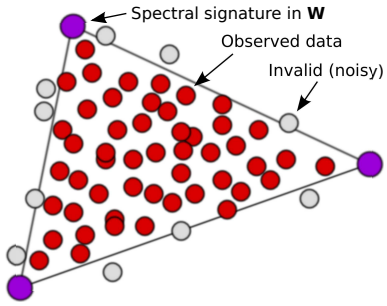
In addition to the non-negativity constraint, the fractional abundances must sum to one[1] in order to maintain proper interpretation. The constraints imposed on the matrix factorization can thus be expressed as

$$w_{mk} \geq 0,\ h_{kn} \geq 0,\ \text{and}\ \sum_{k=1}^{K} h_{k,n} = 1. \tag{2}$$

The non-negativity and sum-to-one assumption of $\boldsymbol{H}$ implies a multidimensional simplex structure of the modeled data, where the vertices denote the endmembers. The simplex is illustrated for two endmembers in one dimension in Figure 2(a) and for three endmembers in two dimensions in Figure 2(b).



(a) Simplex with two endmembers in one dimension



(b) Simplex with three endmembers in two dimensions.

Fig. 2: Illustration of endmembers as the vertices of a simplex. The two illustrations show how the points violating the constraints are located outside the simplices formed by the endmembers (purple circles).

The data might not span the entire simplex due to lack of mixing of the constituents. For food applications, pure endmembers appear rarely as the observed pixels are almost always a mix of constituents. This means the observed data will tend to concentrate around the center of the simplex and very few samples can be expected at the vertices. The data acquisition is further subject to additive noise and thus the simplex structure will not be perfect. Fig-

[1] In the literature, the constraints that abundances must sum to one is sometimes refereed to as an *additivity constraint*.

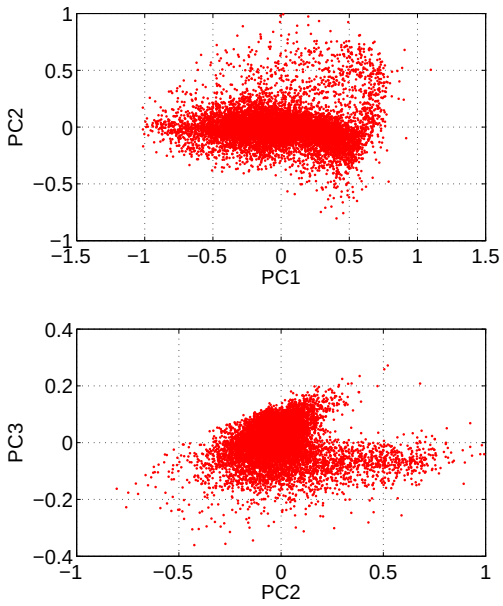


Fig. 3: Scatterplot of first and second principal component of the preprocessed wheat kernel hyperspectral data. The preprocessing pipeline used is described in Section 3.

ure 3 illustrates the simplex structure of wheat kernel image pixels.

Different approaches can be used to unmix the hyperspectral image data. One approach is to analyse the data using convex geometrical methods. Plaza et al. [14] gives a comparative analysis of a set of endmember extraction methods based on geometrical analysis incl. the N-FINDR method [19], VCA [9], PPI [3] and other manually based endmember selection approaches.

Another approach is based on statistical models of the data. As the acquired data is non-negative by nature, non-negative matrix factorization (NMF) has received wide attention [8,11,12,13]. The basic linear NMF model minimizes the squared error $||\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}||^2$ subject to non-negativity constraint on the elements in $\boldsymbol{W}$ and $\boldsymbol{H}$. These constraints are however rarely sufficient to capture the endmembers in an unmixing problem. Furthermore, the solution is not unique since any solution in which all the data are encapsulated by the endmembers will have the same cost; hence, additional regularization is required. This can be expressed through a regularized cost function,

$$C(\boldsymbol{W},\boldsymbol{H}) = \tfrac{1}{2}||\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}||^2 + \gamma J_w(\boldsymbol{W}) + \beta J_h(\boldsymbol{H}), \quad (3)$$

where $J_w(\boldsymbol{W})$ and $J_h(\boldsymbol{H})$ are regularization terms for the endmembers and fractional abundances respectively.

Sajda et al. [15] present an NMF algorithm with additional constraints on the amplitude of the estimated spectral components, with improved endmember determination. An L_2-norm sparsity prior on both the endmembers $\boldsymbol{W}$ and the fractional abundances $\boldsymbol{H}$ is incorporated by Pauca et al. [13], which also leads to improved estimation of the endmembers. A sparsity prior on the fractional abundances $\boldsymbol{H}$ encourages pure spectra among the observed pixels, but in analysis of food data this is rarely the case. Thus, the sparsity prior might not be useful in food applications.

A different approach is to incorporate a regularization based on the volume of the simplex spanned by the estimated endmembers. This encourages a decomposition in which the estimated endmembers lie closer to the data. The volume of a simplex with K vertices $\{\boldsymbol{w}_{:1},\ldots,\boldsymbol{w}_{:K}\}$[2] and $K < M$ is given by (denoted *vol*)

$$J_w^{\text{vol}}(\boldsymbol{W}) = \frac{1}{K!}\sqrt{\det(\tilde{\boldsymbol{W}}^\top \tilde{\boldsymbol{W}})}, \quad (4)$$

where the column vectors of the matrix $\tilde{\boldsymbol{W}} = [\boldsymbol{w}_{:1}-\boldsymbol{w}_{:\rho}\ldots\boldsymbol{w}_{:\rho-1}-\boldsymbol{w}_{:\rho}, \boldsymbol{w}_{:\rho+1}-\boldsymbol{w}_{:\rho}\ldots\boldsymbol{w}_{:K}-\boldsymbol{w}_{:\rho}]$

[2] In the notation used in this paper, matrices and vectors are denoted by capital and lower case bold letters respectively. Two subscripts denotes a sub matrix or sub vector with the corresponding rows and columns, where a colon denotes all indices, and $\tilde{m}$ denotes all indices except m. For example, $\boldsymbol{w}_{:k}$ denotes the kth column of $\boldsymbol{W}$ and $\boldsymbol{w}_{m\tilde{k}}$ denotes the mth row of $\boldsymbol{W}$ with the kth element removed. A single element of the matrix $\boldsymbol{W}$ is denoted by w_{mk}.

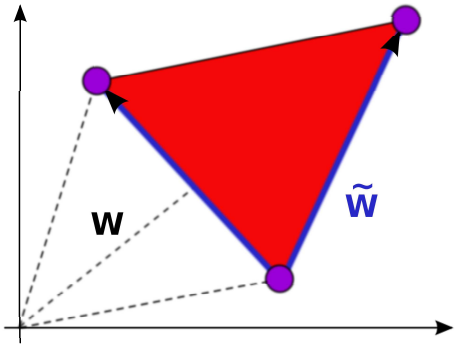


Fig. 4: The volume of a K=3 vertex simplex in two dimensions, illustrating how the volume is computed based on relative endmember vectors.

point to the vertices of the simplex from an arbitrarily selected vertex, $\boldsymbol{w}_\rho$. For $K = M$ Eq. (4) reduces to $J_w^{\text{vol}}(\boldsymbol{W}) = \frac{1}{K!} \det(\tilde{\boldsymbol{W}})$. The case of three endmembers in two dimensions is shown in Figure 4.

Several authors have proposed algorithms for spectral unmixing that employ different volume based regularizations. Miao and Qi [7] present an NMF algorithm that incorporates the following volume regularization (denoted *minimum volume constraint* or *mvc*),

$$J_w^{\text{mvc}}(\boldsymbol{W}) \propto \det^2\big(\boldsymbol{C} + \boldsymbol{B}\boldsymbol{U}_x^\top(\boldsymbol{W} - \boldsymbol{\mu}\mathbf{1}^\top)\big), \quad (5)$$

where $\boldsymbol{U}_x$ is a projection matrix defined as the $K-1$ first principal components of the observed data, $\boldsymbol{X}$, $\boldsymbol{\mu}$ denotes the sample mean of $\boldsymbol{X}$, $\boldsymbol{B} = \left[\begin{smallmatrix} \mathbf{1}^\top \\ \boldsymbol{O} \end{smallmatrix}\right]$ and $\boldsymbol{C} = \left[\begin{smallmatrix} \mathbf{0}^\top \\ \boldsymbol{I} \end{smallmatrix}\right]$. This volume regularization captures the volume of the centered simplex in the subspace defined by the $K-1$ first principal eigenvectors and thus incorporates a noise reduction. This approach is quite intuitive, but due to its dependency on the observed data in $\boldsymbol{U}_x$ it does not have the interpretation as a prior in a Bayesian framework.

Schachtner et al. [16] propose a different volume regularization approach based on the squared volume of the parallelepiped spanned by the endmembers and origo, and they propose an optimization method based on the NMF multiplicative update framework. The regularization term can be expressed as (denoted *parallelepiped* or *pp*)

$$J_w^{\text{pp}}(\boldsymbol{W}) = \det(\boldsymbol{W}^\top\boldsymbol{W}), \quad (6)$$

and can be seen as a surrogate to Eq. (4), where the absolute vectors $\boldsymbol{W}$ are used instead of the simplex spanning vectors $\tilde{\boldsymbol{W}}$. This measure, however, is sensitive to the location of the data simplex as opposed to the simplex volume in Eq. (4). This can potentially lead to movement toward origo when minimizing the volume. Since the regularization is expressed in terms of a squared volume, large volumes will be penalized relatively stronger than small volumes.

Common for the three regularization terms based on the determinant is that they measure volume. Hence, they will tend to shrink the volume, but when the regularization is strong, the K-dimensional volume will collapse to a $(K-1)$-dimensional subspace of which the volume becomes zero.

Another approach, which is not based on a determinant criterion, is to form an approximate volume regularization based on euclidian distance measures. In the ICE algorithm, Berman et al. [2] implement a simplex volume measure as the sum of squared distances between all the vertices of the simplex. Equivalently, we incorporate a measure based on the sum of squared distances from the vertices to the centroid shown in Figure 5 and given by (denoted *dist*),

$$J_w^{\text{dist}}(\boldsymbol{W}) = \sum_{m=1}^{M} \boldsymbol{w}_{m:}\left(\boldsymbol{I} - \tfrac{1}{K}\mathbf{1}\mathbf{1}^\top\right)\boldsymbol{w}_{m:}^\top \quad (7)$$

$$= \sum_{k=1}^{K} \left\| \boldsymbol{w}_{:k} - \frac{1}{K}\sum_{k=1}^{K} \boldsymbol{w}_{:k} \right\|_2^2. \quad (8)$$

This regularization term is not sensitive to the location of the simplex as the *pp* regularization is. With a large regularization, this measure will not collapse the simplex onto a lower dimensional subspace, but will shrink the simplex from each vertex towards the centroid.

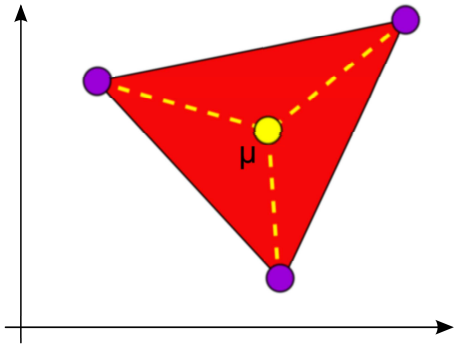


Fig. 5: The *dist* estimates the volume of a simplex based on the vertices euclidian distance to the mean center.

It further has the desirable property of being computationally inexpensive, as it does not require the computation of a determinant.

The *vol*, *pp*, and *dist* regularization terms can be given an interpretation as priors in a Bayesian framework, as they do not depend on the observed data. This is further discussed in the next section.

Common to the volume constrained spectral unmixing methods we have discussed so far is that they specify a regularized cost function and solve for the endmembers by numerical optimization. A different approach for hyperspectral unmixing is to build a probabilistic model and treat endmember extraction as a Bayesian inference problem. This requires the definition of an appropriate likelihood function and priors for the endmembers $\boldsymbol{W}$ and fractional abundances $\boldsymbol{H}$, which is closely related to the choice of cost function and regularization terms. The Bayesian approach further has the advantage of providing credible intervals in addition to an improved estimate of the endmembers and fractional abundances in the form of posterior mean estimates and thereby allowing for a more enhanced analysis.

Previous work on Bayesian spectral unmixing include Ochs et al. [10], who introduce a Bayesian spectral decomposition (BSD) algorithm based on an atomic prior. Moussaoui et al. [8] present a Bayesian method for separating non-negative mixtures of NIR data based on a hybrid Gibbs-Metropolis-Hastings sampling procedure. Schmidt and Laurberg [18] present a Bayesian NMF based on an exponential sparsity prior.

In this paper, which extends our previous work [1], we present a Bayesian spectral unmixing method with a volume prior for unmixing hyperspectral images. The method embodies three different volume priors related to the *vol*, *pp*, and *dist* measures in Eq. (4), (6) and (7). The method incorporates non-negativity constraints for the spectra as well as non-negativity and sum-to-one constraints for the fractional abundances. For model inference, we present a Markov chain Monte Carlo (MCMC) sampling procedure. The details of the method are described in the next section and we refer to the method as *BayesNMF-Vol*.

For comparison to our proposed Bayesian approach, we have implemented optimization based endmember extraction methods using the *mvc*, *pp*, and *dist* volume regularization terms in Eq. (5), (6) and (7). Our implementation is similar to the methods presented in previous work, but is based on a projected gradient NMF framework denoted as *NMF-Vol*. The sum-to-one constraint for the fractional abundances is not included in the work of Schachtner et al. [16] and is implemented as a soft constraint by Miao and Qi [7] in their algorithm. This leads to a trade-off between describing the observed data and respecting the sum-to-one constraint. In our *NMF-Vol* framework the sum-to-one constraint is implemented using with a variable substitution approach, which guarantees that the constraint is fulfilled. We evaluate both the *NMF-Vol* and *BayesNMF-Vol* methods on synthetic an real hyperspectral image data of foods[3].

In Section 2 we present our Bayesian volume constrained NMF model as well as the MCMC sampling procedure. The synthetic and real

[3] Both algorithms are available as a `Matlab` toolbox for download at www.ToBePosted.com

data used in the experiments are described in Section 3 with a description of the pre-processing procedure used. Finally, in Section 4 we evaluate how the proposed method can extract the true endmembers and corresponding fractional abundances, and compare with existing methods.

2 Bayesian NMF with volume prior

The spectral unmixing methods presented in the previous section (except for *mvc*) can be given a Bayesian interpretation: They can be seen as maximum a posteriori (MAP) estimators. The data fit term $\frac{1}{2}||\boldsymbol{X} - \boldsymbol{W}\boldsymbol{H}||^2$ corresponds to a Gaussian likelihood, and the regularization terms, $\gamma J_w(\boldsymbol{W})$ and $\beta J_h(\boldsymbol{H})$ correspond to priors over the parameters $\boldsymbol{W}$ and $\boldsymbol{H}$. Using a Bayesian approach, we are not limited to computing point estimates, such as the MAP estimator, but we can compute the full posterior distribution of the parameters. This can then be used to estimate the parameters and their credible intervals. When we are ultimately interested in a point estimate of the parameters, better estimators than the MAP can be computed, such as the posterior mean or median, which are optimal under squared and linear loss respectively.

2.1 Motivating the Bayesian approach

To further motivate the use of Bayesian methods for spectral unmixing, we consider a very simple one-dimensional example, which clearly demonstrates the differences between the MAP and the Bayesian approach. Consider the case, where we have N measurements, x_n,[4] which are points on the non-negative real line (see Figure 6). We now wish to model these data using the non-negative linear factorization from

[4] For simplicity, we use a slightly different notation in this section.

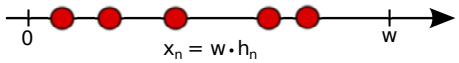


Fig. 6: One-dimensional linear factorization. Data x_n (circles) are points on the non-negative real line. Since $h_n \in [0, 1]$, w must be greater than the maximum data point.

Eq. (1), which in this case can be written as $x_n = w \cdot h_n + \epsilon_n$. For simplicity, we consider the noise free situation, $\epsilon_n = 0$, which results in a likelihood that requires all data points to be modeled exactly,

$$p(\boldsymbol{x}|w, \boldsymbol{h}) = \prod_{i=n}^{N} \delta(x_n - w \cdot h_n). \tag{9}$$

The prior over w is chosen as a flat uninformative (improper) distribution over the non-negative real numbers, which can be thought of as a uniform distribution between zero and infinity,

$$p(w) = \lim_{u_w \to \infty} \frac{1}{u_w} \mathbb{I}\left[0 \leq w \leq u_w\right] \tag{10}$$

$$\propto \mathbb{I}\left[w \geq 0\right]. \tag{11}$$

Here, $\mathbb{I}\left[\cdot\right]$ denotes an indicator function, which has the value one when its argument is true and zero otherwise. The prior for $\boldsymbol{h}$ is chosen as uniform between zero and one,

$$p(\boldsymbol{h}) = \prod_{n=1}^{N} \mathbb{I}\left[0 \leq h_n \leq 1\right]. \tag{12}$$

Our intuition about this model is that w will take some value greater than the maximum data point, and h_n will indicate the fractional distance at which x_n lies. The prior on w captures our ignorance about the location of w, and the prior on h_n states that it corresponds to a proper fraction between zero and one. Data points can be generated from the model by first selecting a random w from the prior, and then for each data point selecting a random h_n between zero and one. The generated data, x_n, will be uniformly distributed between 0 and w, and in the limit of infinitely many

data points, intuition says that w can consistently be estimated as the maximum of the observed data points.

Using Bayes' rule and collecting multiplicative constants, the posterior density of w and $\boldsymbol{h}$ is given by

$$p(w, \boldsymbol{h}|\boldsymbol{x}) = \frac{1}{Z} \prod_{n=1}^{N} \delta(x_n - w \cdot h_n) \times \mathbb{I}\left[0 \leq w\right] \prod_{n=1}^{N} \mathbb{I}\left[0 \leq h_n \leq 1\right], \quad (13)$$

where Z is a normalization constant. It is clear from the posterior, that a MAP estimate of w and $\boldsymbol{h}$ can be found by choosing any w greater than the maximum data value, and then choosing $h_n = \frac{x_n}{w}$. Any solution in which the range $[0, w]$ encapsulates the data, has the same posterior probability density, and the MAP estimate is thus not uniquely defined. For that reason, the MAP estimate is not particularly meaningful in this example; however, this is not because the model is ill defined. The posterior has a ridge of constant probability density, but in the Bayesian approach we are interested in probability mass rather than density. Insight can be gained by looking at the posterior marginals, which are found by integrating over the posterior density. This can be done analytically in this simple example, but requires more elaborate methods such as Markov chain Monte Carlo in the general multidimensional matrix factorization case.

The marginal density of w is given by

$$p(w|\boldsymbol{x}) = \int_{\mathbb{R}^N} p(w, \boldsymbol{h}|\boldsymbol{x}) \mathrm{d}\boldsymbol{h} \quad (14)$$

$$\propto w^{-N} \cdot \mathbb{I}\left[x_{\max} \leq w\right], \quad (15)$$

where $x_{\max} = \max_n(x_n)$. The posterior marginal captures the intuition that w must be greater than the maximum data point, but moreover, it exhibits a polynomial decay, and thus its mass is concentrated in the region close to $x_{\max}$. In the limit $N \to \infty$, the marginal posterior will be infinitely peaked at $x_{\max}$. Similarly, the posterior marginal of h_n is given by

$$p(h_n|\boldsymbol{x}) = \int_{\mathbb{R}^N} p(w, \boldsymbol{h}|\boldsymbol{x}) \mathrm{d}\boldsymbol{h}_{\tilde{n}} \mathrm{d}w \quad (16)$$

$$\propto h_n^{N-2} \cdot \mathbb{I}\left[0 \leq h_n \leq \frac{x_n}{x_{\max}}\right], \quad (17)$$

which in the limit $N \to \infty$ is infinitely peaked at $\frac{x_n}{x_{\max}}$.

In the multidimensional hyperspectral unmixing problem discussed in this paper, data vectors are modeled as lying inside a K-simplex, and the objective is to identify the endmembers (the vertices of the simplex). With non-informative priors, the MAP estimate for this problem is not unique, as discussed previously, since any simplex which encapsulates the data vectors is a MAP solution, analogous to the simple example above. Using an informative prior, such as the volume priors discussed previously, will encourage the simplex to be small, and thus overcome the problem of a non-unique MAP solution. As the example above suggests, when doing full Bayesian inference the non-uniqueness of the MAP solution is not an issue of concern — even when using non-informative priors. The reason is that although the maximum of the posterior is not unique, the posterior density itself is uniquely determined.

In addition to making the MAP estimate well determined, the different volume priors suggested in the literature also serve another purpose. Real data from hyperspectral imaging problems do not in general exactly obey the linear mixing property, and there might be outlying data points, etc. Thus, the regularization parameter γ in the volume prior can be used to push the algorithm towards a good solution. In the following we derive a Bayesian inference procedure for hyperspectral unmixing, which incorporates three different volume priors.

2.2 Model

We model the joint probability distribution of the endmembers, $\boldsymbol{W}$, and the fractional abundances, $\boldsymbol{H}$, as well as the noise, conditioned on the observed data, $\boldsymbol{X}$, and a set of model hyper-parameters, $\mathcal{H}$. To this end we choose a suitable noise model as well as reasonable prior distributions over all parameters in the model.

2.2.1 Noise model

We model the noise, $\boldsymbol{\epsilon}$, as independent and identically distributed white Gaussian noise, which gives rise to the following likelihood function,

$$p(\boldsymbol{X}|\boldsymbol{W},\boldsymbol{H},\sigma^2) = \prod_{n=1}^{N}\prod_{m=1}^{M}\mathcal{N}\left(x_{mn}|\boldsymbol{w}_{m:}\boldsymbol{h}_{:n},\sigma^2\right), \quad (18)$$

where $\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(\frac{(x-\mu)^2}{-2\sigma^2}\right)$ is the Gaussian probability density function. Note that the negative logarithm of the likelihood function corresponds to the squared error criterion which is the first term in Eq. (3).

The likelihood has a single parameter, the noise variance, σ^2, for which we choose a conjugate prior, i.e., an inverse-Gamma distribution,

$$p(\sigma^2|\alpha,\beta) = \mathcal{IG}(\sigma^2|\alpha,\beta) \quad (19)$$

$$= \frac{\beta^\alpha}{\Gamma(\alpha)}\left(\frac{1}{\sigma^2}\right)^{\alpha+1}\exp\left(-\frac{\beta}{\sigma^2}\right). \quad (20)$$

2.2.2 Model of fractional abundances

For the fractional abundances, $\boldsymbol{H}$, the prior must enforce non-negativity as well as the constraint that the abundances for each pixel must sum to unity. We choose a uniform prior on the unit simplex,

$$p(\boldsymbol{H}) \propto \prod_{n=1}^{N}\mathbb{I}\left[\sum_{k=1}^{K}h_{kn}=1\right]\prod_{k=1}^{K}\mathbb{I}\left[h_{kn}\geq 0\right],$$

which is arguably the simplest and most noninformative prior that expresses these constraints.

2.2.3 Model of endmembers

We choose a prior distribution for the endmembers, that encourages the simplex spanned by the estimated endmembers to be small, and which includes the constraint that each element in the endmember matrix must be non-negative,

$$p(\boldsymbol{W}|\gamma) \propto e^{-\gamma J_w(\boldsymbol{W})}\prod_{m=1}^{M}\prod_{k=1}^{K}\mathbb{I}\left[w_{mk}\geq 0\right]. \quad (22)$$

The reason for choosing a prior proportional to $e^{-\gamma J_w(\boldsymbol{W})}$ is that the negative logarithm of the prior then corresponds to the regularization term, $\gamma J_w(\boldsymbol{W})$, in the cost function defined in Eq. (3). Thus, there is a direct parallel between the methods discussed in the introduction and our Bayesian probabilistic model. Specifically, we consider three different volume measures,

$$J_w^{\mathrm{pp}}(\boldsymbol{W}) = \det(\boldsymbol{W}^\top\boldsymbol{W}), \quad (23)$$

$$J_w^{\mathrm{sv}}(\boldsymbol{W}) = \det(\tilde{\boldsymbol{W}}^\top\tilde{\boldsymbol{W}}), \text{ and} \quad (24)$$

$$J_w^{\mathrm{dist}}(\boldsymbol{W}) = \sum_{k=1}^{K}\left|\left|\boldsymbol{w}_{:k}-\frac{1}{K}\sum_{k=1}^{K}\boldsymbol{w}_{:k}\right|\right|_2^2. \quad (25)$$

$J_w^{\mathrm{pp}}(\boldsymbol{W})$ measures the squared volume of the parallelepiped defined by the endmembers and the origin, $J_w^{\mathrm{sv}}(\boldsymbol{W})$ is $(K!)^2$ times the squared volume of the simplex spanned by the endmembers, and $J_w^{\mathrm{dist}}(\boldsymbol{W})$ measures the sum of squared distances from the endmembers to their centroid. We have chosen these three measures of the simplex volume, because they can all be written as quadratic polynomials when considered as functions of a single element w_{mk}, which is easy to see, since the determinant is linear in its argument. For this reason, the prior has the form of a truncated Gaussian,

when considered as a function of a single element w_{mk}. This makes it possible to derive a direct Gibbs sampling procedure, as we describe in the next section. The parameter γ, which is common for these three volume priors, determines the strength of the volume penalty. In this paper, we study how the strength of volume prior influences the solutions and in our experiments we generate sets of solutions for a range of values of γ as an explorative analysis. If a single solution is of interest, γ can also be modeled hierarchically, or a good value for γ can be found by cross validation or Bayesian model comparison.

2.2.4 Posterior

Using Bayes' rule, the posterior is given by

$$p(\boldsymbol{W}, \boldsymbol{H}|\boldsymbol{X}, \mathcal{H}) = \frac{p(\boldsymbol{X}|\boldsymbol{W}, \boldsymbol{H}, \sigma^2)p(\boldsymbol{H})p(\boldsymbol{W}|\gamma)p(\sigma^2|\alpha, \beta)}{p(\boldsymbol{X})}, \tag{26}$$

where $\mathcal{H} = \{\alpha, \beta, \gamma\}$ are hyperparameters. In the following we let $\mathcal{P} = \{\boldsymbol{W}, \boldsymbol{H}, \sigma^2\}$ denote the parameters of the model.

2.3 Gibbs sampler

A Gibbs sampling procedure [4] can be used to infer the posterior distribution of the parameters of the model, $\mathcal{P}$. In Gibbs sampling, we sequentially draw samples from the posterior of each parameter, conditioned on all other parameters. It can be shown that the sequence of samples computed constitute a Markov chain for which the stationary distribution is the posterior in which we are interested. Due to our choice of priors, we can sample from all conditional distributions directly using standard methods, which obviates slower sampling procedures such as rejection sampling.

2.3.1 Sampling the noise variance

Because we have chosen a conjugate prior for the noise variance, σ^2, its conditional distribution has the same functional form as the prior: It is an inverse-Gamma,

$$p(\sigma^2|\boldsymbol{X}, \mathcal{P}\backslash\sigma^2) = \mathcal{IG}(\sigma^2|\bar{\alpha}, \bar{\beta}), \tag{27}$$

where the parameters are given by

$$\bar{\alpha} = \alpha + \tfrac{1}{2}NM, \tag{28}$$

$$\bar{\beta} = \beta + \tfrac{1}{2}\sum_{m=1}^{M}\sum_{n=1}^{N}\left(x_{mn} - \boldsymbol{w}_{m:}\boldsymbol{h}_{:n}\right)^2. \tag{29}$$

Samples from this distribution can be generated using standard methods.

2.3.2 Sampling fractional abundances

The conditional density of the fractional abundances, $\boldsymbol{H}$, arises from the product of the Gaussian likelihood and the uniform prior on the unit simplex and it hence has the form of a Gaussian constrained to lie on the unit simplex,

$$p(\boldsymbol{h}_{:n}|\boldsymbol{X}, \mathcal{P}\backslash\boldsymbol{h}_{:n}) \propto \mathcal{N}(\boldsymbol{h}_{:n}|\bar{\boldsymbol{\mu}}_n, \bar{\boldsymbol{\Sigma}}_n) \times \prod_{n=1}^{N}\mathbb{I}\left[\sum_{k=1}^{K}h_{kn} = 1\right]\prod_{k=1}^{K}\mathbb{I}\left[h_{kn} \geq 0\right], \tag{30}$$

and its posterior conditional parameters are given by

$$\bar{\boldsymbol{\mu}}_n = (\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{x}_{:n}, \tag{31}$$

$$\bar{\boldsymbol{\Sigma}}_n = \sigma^2(\boldsymbol{W}^\top\boldsymbol{W})^{-1}. \tag{32}$$

Samples from the constrained Gaussian density can be generated using using the method described by Schmidt [17].

2.3.3 Sampling endmembers

The conditional distribution of the endmembers, $\boldsymbol{W}$, arises from the product of the Gaussian likelihood and the volume penalizing prior. As noted earlier, the conditional prior has the form of a truncated Gaussian,

$$p(w_{mk}|\mathcal{P}\backslash w_{mk}) \propto \mathcal{N}(w_{mk}|m_{mk}, s^2_{mk})\mathbb{I}\left[w_{mk} \geq 0\right], \quad (33)$$

assuming $\boldsymbol{W}$ satisfies the non-negativity constraint. The values of m_{km} and s^2_{km} depend on the choice of $J_w(\boldsymbol{W})$. Using this, the posterior conditional is a truncated Gaussian

$$p(w_{mk}|\boldsymbol{X}, \mathcal{P}\backslash w_{mk}) \propto \mathcal{N}(w_{mk}|\bar{\mu}_{mk}, \bar{\sigma}^2_{mk})\mathbb{I}\left[w_{mk} \geq 0\right], \quad (34)$$

with parameters given by

$$\bar{\sigma}^{-2}_{mk} = s^{-2}_{mk} + (\boldsymbol{h}_{k:}\boldsymbol{h}^\top_{k:})\sigma^{-2} \quad (35)$$

$$\bar{\mu}_{mk} = \bar{\sigma}^2_{mk}\Big(m_{mk}s^{-2}_{mk}, \quad (36)$$

$$+ \left(\boldsymbol{x}_{m:}\boldsymbol{h}^\top_{k:} - \boldsymbol{w}_{m\tilde{k}}\boldsymbol{H}_{\tilde{k}:}\boldsymbol{h}^\top_{k:}\right)\sigma^{-2}\Big). \quad (37)$$

Samples from this distribution can be generated using the methods described by Geweke [5] and Schmidt [17]. What is left is to derive expressions for m_{mk} and s^2_{mk} for each of the three volume measures in Eq. (23–25).

For the determinant based measures, we use the following expression for the determinant of a symmetric matrix,

$$\det\begin{bmatrix} a & \boldsymbol{b}^\top \\ \boldsymbol{b} & \boldsymbol{C} \end{bmatrix} = a\det\left(\boldsymbol{C} - \tfrac{1}{a}\boldsymbol{b}\boldsymbol{b}^\top\right) \quad (38)$$

$$= a\det(\boldsymbol{C})\left(1 - \tfrac{1}{a}\boldsymbol{b}^\top\boldsymbol{C}^{-1}\boldsymbol{b}\right) \quad (39)$$

$$= a\det(\boldsymbol{C}) - \boldsymbol{b}^\top\mathrm{adj}(\boldsymbol{C})\boldsymbol{b}, \quad (40)$$

where (38) is the expression for the determinant of a block matrix, (39) follows from the matrix determinant lemma, and in (40) we have used the definition of the matrix adjugate. Using this, for $-\gamma J^{\mathrm{pp}}_w$ we arrive at

$$s^{-2}_{mk} = \gamma\left(d_{\tilde{k}\tilde{k}} - \boldsymbol{w}_{m\tilde{k}}\boldsymbol{A}_{\tilde{k}\tilde{k}}\boldsymbol{w}^\top_{m\tilde{k}}\right), \quad (41)$$

$$m_{mk} = s^2_{mk}\gamma\boldsymbol{w}_{m\tilde{k}}\boldsymbol{A}_{\tilde{k}\tilde{k}}\boldsymbol{W}^\top_{\tilde{m}\tilde{k}}\boldsymbol{w}_{\tilde{m}k}, \quad (42)$$

where $d_{\tilde{k}\tilde{k}}$ and $\boldsymbol{A}_{\tilde{k}\tilde{k}}$ are the determinant and adjugate of $\boldsymbol{W}^\top_{:\tilde{k}}\boldsymbol{W}_{:\tilde{k}}$ respectively, and for $-\gamma J^{\mathrm{sv}}_w$ we get

$$s^{-2}_{mk} = \gamma\left(\tilde{d}_{\tilde{k}\tilde{k}} - \tilde{\boldsymbol{w}}_{m\tilde{k}}\tilde{\boldsymbol{A}}_{\tilde{k}\tilde{k}}\tilde{\boldsymbol{w}}^\top_{m\tilde{k}}\right), \quad (43)$$

$$m_{mk} = w_{m\rho} + s^2_{mk}\gamma\tilde{\boldsymbol{w}}_{m\tilde{k}}\tilde{\boldsymbol{A}}_{\tilde{k}\tilde{k}}\tilde{\boldsymbol{W}}^\top_{\tilde{m}\tilde{k}}\tilde{\boldsymbol{w}}_{\tilde{m}k}, \quad (44)$$

where $\rho \neq k$, and $\tilde{d}_{\tilde{k}\tilde{k}}$ and $\tilde{\boldsymbol{A}}_{\tilde{k}\tilde{k}}$ are the determinant and adjugate of $\tilde{\boldsymbol{W}}^\top_{:\tilde{k}}\tilde{\boldsymbol{W}}_{:\tilde{k}}$ respectively. For $-\gamma J^{\mathrm{dist}}_w$ we get

$$s^{-2}_{mk} = \gamma\tfrac{N-1}{N}, \; m_{mk} = \tfrac{1}{N-1}\sum_{k'\neq k} w_{mk'}. \quad (45)$$

Details of the derivations can be found in a technical note at www.ToBePosted.com.

3 Data acquisition and pre-processing

The goal of this work is to develop useful methods for unmixing hyperspectral images of wheat kernels. A wheat kernel consists of many different constituents, where the majority is *water*, *starch*, *protein* and *oil*. Hence we expect to extract $4-5$ endmembers including background in our analysis. For the acquisition of the image, a hyperspectral NIR line scan camera from Headwall is used from 900-1700nm in 165 bands. Hence each acquired image becomes a 3-way tensor of size $320 \times lines \times 165$. Two hyperspectral images of 14 wheat kernels are acquired on both front and back side, depicted in Figure 7.

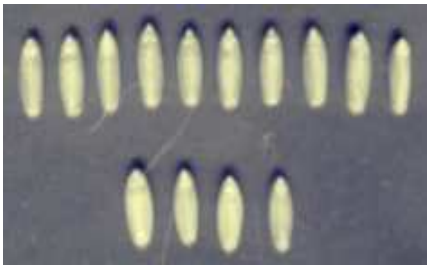


Fig. 7: The raw acquired image of the front side of the wheat kernels in pseudo colors.

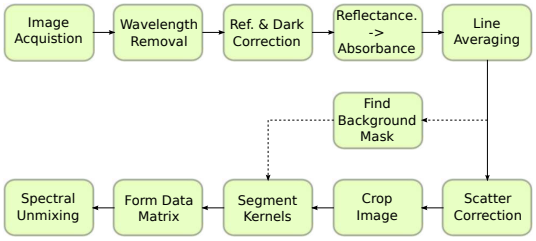


Fig. 8: Flowchart of pre-processing pipeline.

Prior to the image data being subjected to our unmixing algorithm a series of pre-processing steps are carried out shown in Figure 8.

Initially the peripheral spectral range from 900-950nm and 1650 − 1700nm are removed due to the poor signal-to-noise ratio of the camera sensor. Afterwards, the raw image data is compensated for white reference and dark current background spectra to remove light source characteristics and sensor offset. In order to suppress noise, each line was scanned twice and averaged.

The light scattering effects induced in the observed data are in our case compensated in two steps. First each spectra is converted to absorbance by a negative log-transform to obey Lambert-Beer's law for light transmission. Secondly a scatter correction step aligns the spectra as shown in Figure 9. As sparse spectral peaks are not prominent in NIR data, low order scatter correction is applied as the residual from a first order polynomial fit. This approach still preserves the simplex structure as shown in Figure 3.

For our analysis, the individual wheat kernels are extracted or cropped from the images by identifying and removing the pure background pixels from the data set. This segmentation is achieved by discriminating the first principal component applied on the image data depicted in Figure 10. A few background pixels are left around the kernel periphery as illustrated in Figure 11 in order to capture an entire grain kernel. This further allows the background to be identified as a single endmember

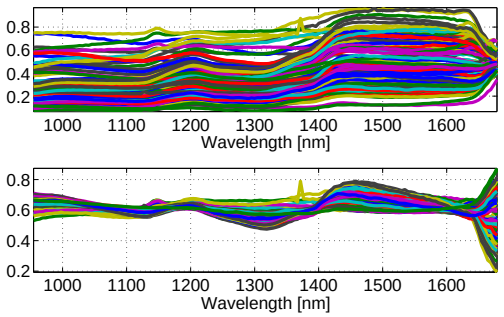


Fig. 9: The observed spectra before and after the affine scatter correction (above and below respectively).

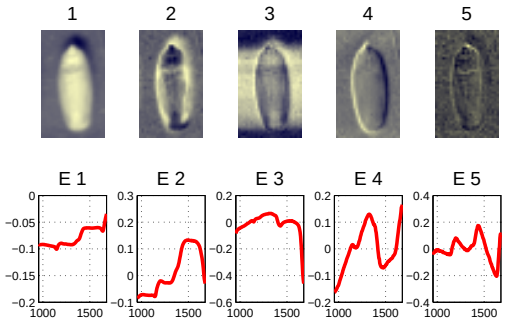


Fig. 10: Principal component images of the wheat kernel data set with corresponding associated PC spectra enumerated E1-E5.

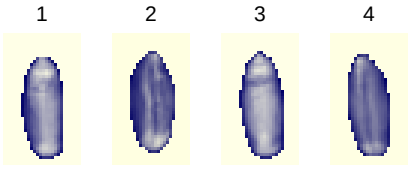
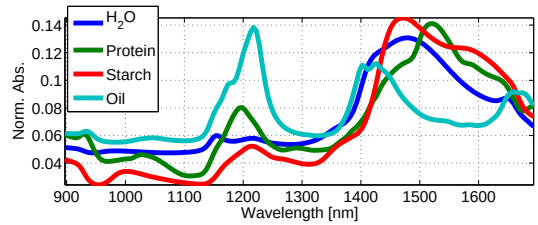


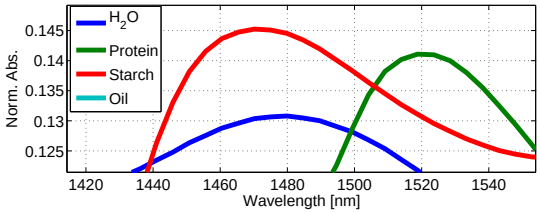
Fig. 11: A few segmented kernels with a little background included in the periphery.

component in our subsequent unmixing. The final hyperspectral image data set is then represented as an unfolded matrix, $\boldsymbol{X} \in \mathbb{R}^{M \times N}$, where each M-dimensional column vector $\boldsymbol{x}_{:n}$ is a pre-processed spectrum of a pixel.

For reference spectra of pure food constituents (protein, starch and oil) are acquired using the same camera system and pre-processed as described. These measurements can act as refer-



(a)



(b) Zoom showing the clear difference in peak location between protein and starch.

Fig. 12: Normalized spectra of pure basic food constituents. The water spectrum may suffer from poor SNR as H_2O has extremely high absorption rates from $1400 - 1700$nm.

ences in evaluating unmixed spectra and are shown in Figure 12. Initially the spectral profiles of protein and starch appear similar in most of the wavelength range. The most prominent spectral difference is around 1450-1550nm as shown in Figure 12(b). Any relocation of the peak can be used to indicate different protein/starch mixtures and hence to interpret unmixing spectral results.

4 Simulations

Our two volume regularized spectral unmixing algorithms, *NMF-Vol* and *BayesNMF-Vol*were evaluated in a set of simulations using synthetic data with different profiles and known labels. Next, our *BayesNMF-Vol* model was applied to real hyperspectral image data of wheat kernels.

4.1 Synthetic data

A set of synthetic datasets with different profiles of noise levels and dimensionality were produced with full mixing, i.e. the generated samples spanned the entire simplex. Several synthetic datasets were created with different amount of endmembers, varying data dimensionality, and noise level. The first experiments have few endmember in a low-dimensional space. The next include a realistic number of endmembers and data dimensions found in real hyperspectral images. Our final experiments are conducted with a small sample size, where the dimensionality of the data is greater than the number of samples. Table 1 lists the different synthetic data set profiles. Each dataset was based on uniform distributed randomly generated endmembers $\mathbf{W}$ used to produce different amounts of fractional abundance samples $\mathbf{H}$.

The samples were generated using the *simplex point picking* procedure leading to a uniform distribution of the generated samples $\mathbf{H}$ over the unit simplex. The procedure exploits the fact that the simplex can be considered a Dirichlet distribution with all parameters set to 1. Initially uniform random samples between $[0; 1]$ are drawn as h_{uni} and processed as $h = -log(h_{uni})$. They are afterwards scaled to $h_{k,n} = h_{k,n}/\sum_{k'} h_{k',n}, \forall k$. Figure 13 illustrates an example of a synthetically generated dataset as a principal component scatterplot[5].

The synthetic datasets were initially analyzed using *NMF-Vol* with three different regularizations; the *mvc* (5), *pp* (6) and *dist* (7) measures. All algorithms were initialized from randomly selected observed data points among $\mathbf{X}$ and a long exponential range of regularization parameter values γ were applied from $[10^{-6} - 10^4]$. Finally, the endmember MAP estimates $\mathbf{W}_{map}$ for the different regularizations were found. Similarly, our *BayesNMF-*

[5] All scatterplots in this paper are presented as subspace projections onto the first and second PC based on the data points.

Table 1: List of synthetic data sets. Boldfaced will be used for illustration.

ID	Endmembers K	Dim. M	Noise Var. σ^2	Samples N
1	3	3	10^{-4}	3000
2	**3**	**3**	$\mathbf{10^{-3}}$	**3000**
3	**3**	**150**	$\mathbf{10^{-4}}$	**3000**
4	**3**	**150**	$\mathbf{10^{-2}}$	**3000**
5	5	150	10^{-4}	5000
6	5	150	10^{-3}	5000
7	**5**	**150**	$\mathbf{10^{-1}}$	**5000**
8	5	150	10^{-4}	50
9	5	150	10^{-3}	50
10	**5**	**150**	$\mathbf{10^{-2}}$	**50**

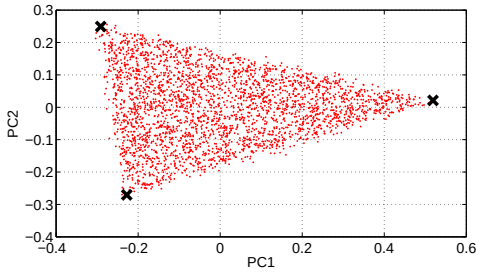


Fig. 13: Scatter plot of the synthetic data for, 1st and 2nd PC. Note the uniform distribution of the samples over the unity simplex. *(Dots denote data points and crosses denote endmembers.)*

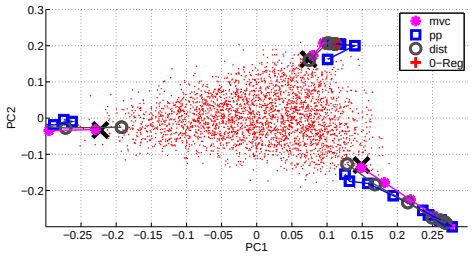
Vol method was applied to the same datasets with a similar range of appropriate regularization. We used the tree different volume priors; *pp* (23), *sv* (24), and *dist* (25). For each regularization level we generated 3000 Gibbs samples and disregarded the initial 2000 samples as burn in to calculate the posterior mean endmember estimates. A more thorough simulation generating 12000 Gibbs samples incl. 2000 burn-in samples were used to produce and evaluate the credible intervals. Due to the light regularization latent in the BayesNMF-Vol algorithm caused by likelihood normalization a non-informative prior of $\gamma = 0$ was sufficient for this latter simulation.

Results of a subset of our experiments (boldface in Table 1) are shown in Figures 14–17 illustrating the converged endmember estimates for both algorithms with all regularization approaches and with interconnecting lines for the different regularization levels. The models successfully capture the data structures with more samples than dimensions and reveal similar performance in endmember extraction $\mathbf{W}$. The *BayesNMF-Vol* model successfully captures the endmembers on par with the three regular volume regularized NMF models. There is, however, a clear difference between the two approaches: With low regularization, the *NMF-Vol* methods find endmembers outside the data simplex, and thus need a suitable regularization to give a reasonable answer. This is especially visible in Figure 14(a). The *BayesNMF-Vol* with low regularization gives solutions close to the true endmembers, and higher regularization tends to shrink the simplex further as desired. A major advantage is the possibility to compute credible intervals for the endmembers, as depicted in Figures 14(c)–17(c). This means the confidence of the estimate can be evaluated. Increasing the regularization encourages smaller volumes, as expected, but is not necessarily required for the Bayesian method. We do expect, however, that it will be useful in real data to counteract noise and outliers.

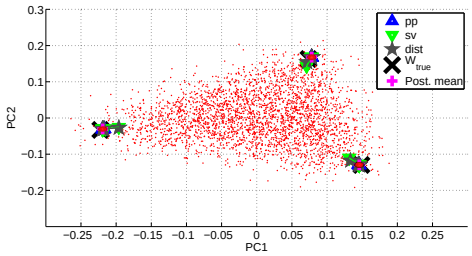
In practice the different regularizations all suppress the noise to enhance the simplex structure in the data. Depending on the level of noise a suitable level of regularization can be found and hence the unmixing performance across the parameters are expected to be comparable.

The data structures with fewer samples than dimension are a bit more difficult to unmix as seen in Figure 17. The estimated endmembers fluctuate more due to the lack of structural representation in the small amount of samples. This is also manifested in the larger variance for the BayesNMF-Vol method in Figure 17(c).

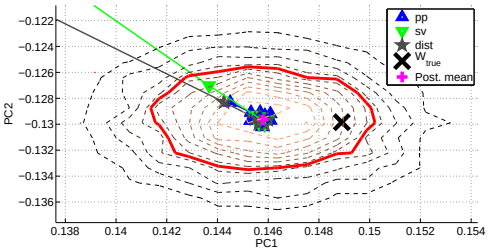
The differences between the determinant and distance based volume regularizations, as discussed in the introduction, is illustrated in Figure 18. With increasing regularization strength, the *dist* prior shrinks toward the center of the



(a) *NMF-Vol* showing MAP estimates.

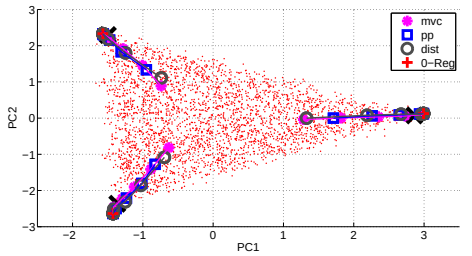


(b) *BayesNMF-Vol* with posterior mean endmember estimates.

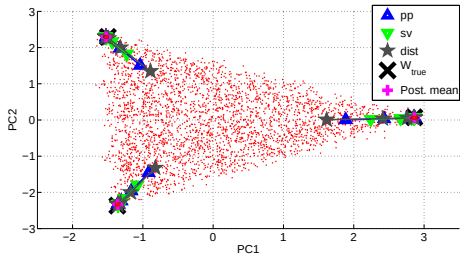


(c) *BayesNMF-Vol* zoomed with post. mean endmember estimates. Bold red line is 90% credible interval.

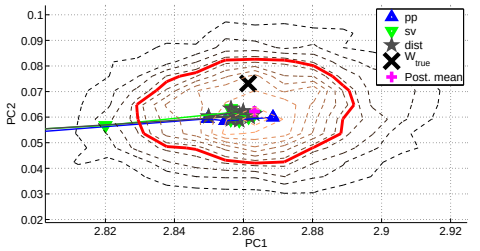
Fig. 14: Scatterplot of unmixing result with parameters $ID = 2, M = 3, K = 3, \sigma^2 = 10^{-3}$. Different regularization levels are shown with interconnecting lines.



(a) *NMF-Vol* showing MAP estimates.



(b) *BayesNMF-Vol* with posterior mean endmember estimates.



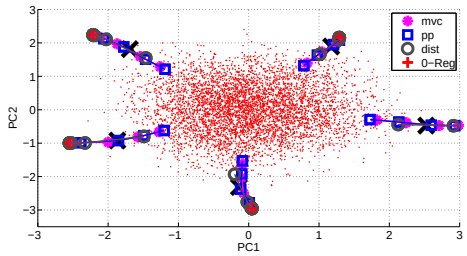
(c) *BayesNMF-Vol* zoomed with post. mean endmember estimates. Bold red line is 90% credible interval.

Fig. 15: Scatterplot of unmixing result with parameters $ID = 4, K = 3, M = 150, \sigma^2 = 10^{-2}$. Different regularization levels are shown with interconnecting lines.

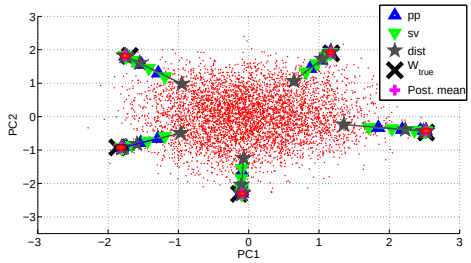
simplex, whereas both the *pp* and the *sv* priors initially shrink the simplex and eventually collapses the simplex onto a line along the principal eigenvector. This result is intuitive, since the collapsed simplex has zero volume, but still extends along a direction that can explain the variance of the data.

4.2 Wheat kernel data

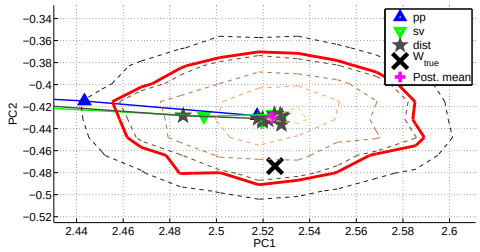
The wheat kernel data comprises of 14 grains, where 4 are selected with front and back side to be used for the unmixing, i.e, a total of 8 kernel images. Based on their biological properties, wheat kernels consists of many different constituents, where the majority is *water*, *starch*, *protein* and *oil*, i.e. we expect to extract $4 - 5$ endmembers incl. background in our analysis.



(a) *NMF-Vol* showing MAP estimates.

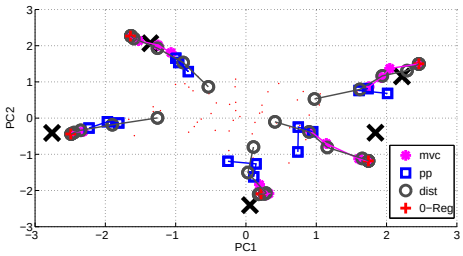


(b) *BayesNMF-Vol* with posterior mean endmember estimates.

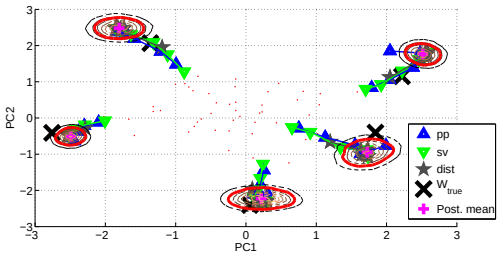


(c) *BayesNMF-Vol* zoomed with post. mean endmember estimates. Bold red line is 90% credible interval.

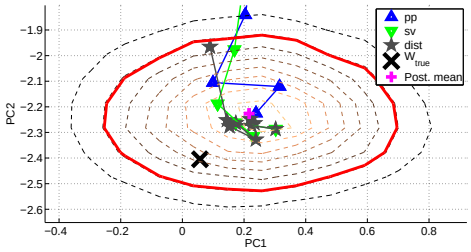
Fig. 16: Scatterplot of unmixing result with parameters $ID = 7, K = 5, M = 150, \sigma^2 = 10^{-1}$. Different regularization levels are shown with interconnecting lines.



(a) *NMF-Vol* showing MAP estimates.



(b) *BayesNMF-Vol* with posterior mean endmember estimates.



(c) *BayesNMF-Vol* zoomed with post. mean endmember estimates. Bold red line is 90% credible interval.

Fig. 17: Scatterplot of unmixing result with parameters $ID = 10, K = 5, M = 150, \sigma^2 = 10^{-2}$. Different regularization levels are shown with interconnecting lines.

A set of reference concentrations for the constituents are unfortunately not available, so we perform a subjective evaluation of the unmixing results. The 4 kernels may suffer from only small variations in the protein level between the pixels, which means that there is a strong correlation with other constituents. In such case the spectral profile of protein is almost impossible to extract. Principal component scatterplots of the wheat kernel data suggest a simplex structure of 3-4 components as illustrated in Figure 19.

The BayesNMF-Vol algorithm was initialized with *pp* regularized *NMF-Vol* endmember estimates in order to avoid unnecessary long burn-in periods. The *pp* prior was also chosen for the unmixing and a suitable prior parameter, γ was found empirically through manual experiments. Figure 20 shows the estimated posterior mean endmembers for $K = 4$ endmembers and 10000 Gibbs samples having disregarded initial 10000 burn-in samples.

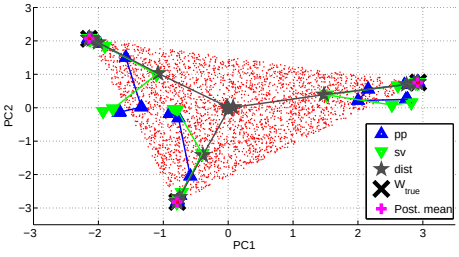


Fig. 18: Scatterplot of unmixing result with parameters $ID = 3, K = 3, M = 150, \sigma^2 = 10^{-4}$.

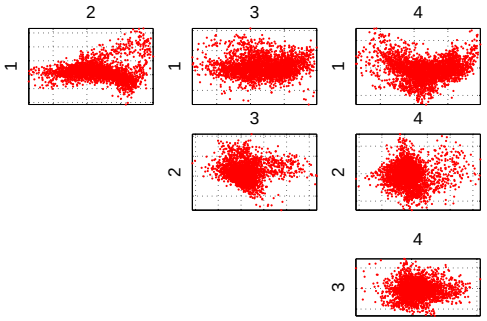
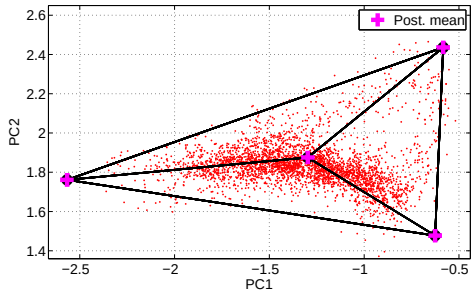


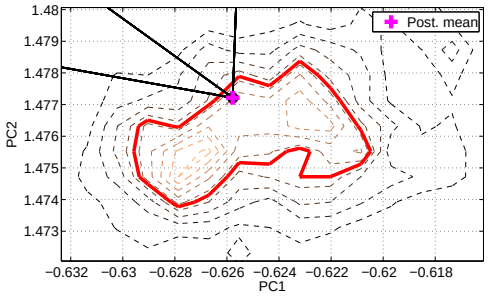
Fig. 19: PCA scatter plot of the wheat kernel data indicating a strong simplex structure in the first 3 PC. This suggests $3-4$ endmembers latent in the data set.

The spectral profiles of the endmembers, illustrated in Figure 21(a), shows how the 2nd and 4th spectra designate the starch and protein content. The most prominent difference is the position of the peak at appr. 1450nm. indicating different mixtures of protein and starch as compared with the reference measurements in Figure 12(b).

Of the 4 kernels (front and backside) a single decomposed wheat kernel is illustrated in Figure 21(b). Initially the background has been extracted by the model as the 1st component. The protein and starch spectral profile have also been identified very clearly along with a spatial distribution in the 2nd and 4th component. Similarly the oil in the germ part can be identified primarily from the spatial distribution. Finally the residual reveals very little structure suggesting a successful decomposition.

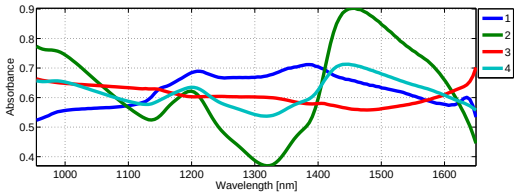


(a) Scatterplot of estimated simplex.

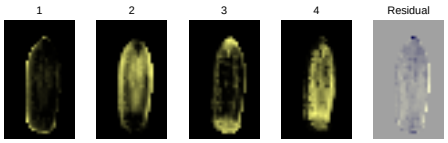


(b) Zoomed, bold red line is 90% credible interval.

Fig. 20: BayesNMF-Vol analysis of wheat kernel data.



(a) Extracted spectral endmembers as the posterior mean.



(b) Image components of an unmixed wheat kernel as the posterior means.

Fig. 21: The 2nd and 4th extracted spectral endmember (green and cyan) are easily identified as starch/protein matrix comparing to figure 12.

4.3 Discussion

All the volume based NMF methods described show light sensitivity to the number of components, K. If K is too large, excessive components are typically split up in two. In contrast component will be merged if K is too small. We have found that the regularization terms / priors based on the volume as computed using the determinant has a sometimes fatal sensitivity to linear dependencies among the estimated endmembers leading to a collapsing volume. This can occur if the number of components K used in the analysis is greater than the true number of components in the data. In such situation, the excess endmember can cause the simplex to collapse and the volume regularization will be rendered inoperative. The euclidian distance based volumes does not suffer from the same issue as it simply shrinks the volume according to the strength of the regularization/ prior. It can also occur that a strong regularizations, i.e. large value of γ, leads to a collapse of the volume, as it is never allowed to expand and capture the data. A useful approach to estimate the true amount of endmembers K is to conduct a regular PCA scatterplot and evaluate the required number of components to capture the simplex structure. Prior knowledge can also be used such as general biological properties for wheat kernels for instance, similar to our analysis in section 4.2.

In addition, our *BayesNMF-Vol* algorithm conducts light regularization in it self on the estimated endmembers as discussed. For a flat prior distribution $p(\mathbf{W})$ with $\gamma = 0$, the BayesNMF-Vol still encourages small volumes and does not necessarily require regularization to give meaningful results. This also means the observed data point may not be encapsulated entirely when using *BayesNMF-Vol* with $\gamma = 0$.

The synthetic data we used were limited to full mixing profiles, where the entire simplex is spanned with generated samples. In many real world datasets this might not always be true, as for our wheat kernel dataset for instance. Future improvement would therefore be to generate synthetic dataset with different mixing profiles to evaluate our methods performance in this area.

Both the *NMF-Vol* and *BayesNMF-Vol* are available as a `Matlab` toolbox with a few examples at http://www.ToBePosted.com.

5 Conclusion

We have proposed a Bayesian method for spectral unmixing, employing a volume based prior suitable for hyperspectral image analysis of foods. Results on synthetic data sets indicate similar or better unmixing performance compared to existing volume regulated NMF models and can further give credible intervals.

In addition we have identified known issues and limitations to our methods and proposed several remedies and approaches to circumvent them. This has further given rise to improvements and future work.

In a concrete food application of decomposing wheat kernels into constituents our methods prove successful and can be used as part of assessment of the quality of foods.

References

1. Arngren, M., Schmidt, M., Larsen, J.: Bayesian nonnegative matrix factorization with volume prior for unmixing of hyperspectral images. In: Machine Learning for Signal Processing, IEEE Workshop on (MLSP) (2009). DOI 10.1109/MLSP.2009.5306262
2. Berman, M., Kiiveri, H., Lagerstrom, R., Ernst, A., Dunne, R., Huntington, J.: ICE: a statistical approach to identifying endmembers in hyperspectral images. IEEE Transactions on Geoscience and Remote Sensing **Vol. 42**(10), 2085–2095 (2004)
3. Boardman, J.W., Kruse, F.A., Green, R.O.: Mapping target signatures via partial unmixing of aviris data. Summaries of JPL Airborne Earth Science Workshop (1995)

4. Geman, S., Geman, D.: Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. Pattern Analysis and Machine Intelligence, IEEE Transactions on **Vol. 6**, 721–741 (1984). DOI 10.1109/TPAMI.1984.4767596
5. Geweke, J.: Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities. Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface pp. 571–578 (1991)
6. Masalmah, Y.M.: Unsupervised unmixing of hyperspectral imagery using the constrained positive matrix factorization. Ph.D. thesis, Computing And Information Science And Engineering, University Of Puerto Rico, Mayagüez Campus (2007)
7. Miao, L., Qi, H.: Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization. Geoscience and Remote Sensing, IEEE Transactions on **Vol. 45**(3), 765–777 (2007)
8. Moussaoui, S., Brie, D., Mohammad-Djafari, A., Carteret, C.: Separation of non-negative mixture of non-negative sources using a bayesian approach and mcmc sampling. IEEE Transactions on Signal Processing **Vol. 54**(11), 4133–4145 (2006)
9. Nascimento, J.M.P., Dias, J.M.B.: Vertex component analysis: A fast algorithm to unmix hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing **Vol. 43**(4), 898–910 (2005)
10. Ochs, M.F., Stoyanova, R.S., Arias-Mendoza, F., Brown, T.R.: A new method for spectral decomposition using a bilinear bayesian approach. Journal of Magnetic Resonance **Vol. 137**(1), 161–176 (1999)
11. Paatero, P., Tapper, U.: Positive matrix factorization: A non-negative factor model with optimal utilization of error estimates of data values. Environmetrics **Vol. 5**(2), 111 – 126 (1994)
12. Parra, L., Spence, C., Sajda, P., Ziehe, A., Müller, K.R.: Unmixing hyperspectral data. Neural Information Processing Systems **Vol. 12** (1999)
13. Pauca, V.P., Piper, J., Plemmons, R.J.: Nonnegative matrix factorization for spectral data analysis. Linear Algebra and Its Applications **Vol. 416**(1), 29–47 (2006)
14. Plaza, A., Martínez, P., Pérez, R., Plaza, J.: A quantitative and comparative analysis of endmember extraction algorithms from hyperspectral data. IEEE Transactions on Geoscience and Remote Sensing **Vol. 42**(3) (2004)
15. Sajda, P., Du, S., Parra, L.: Recovery of constituent spectra using non-negative matrix factorization. Proceedings of the SPIE - The International Society for Optical Engineering **Vol. 5207**(1), 321–31 (2003)
16. Schachtner, R., Pöppel, G., Tomé, A.M., Lang, E.W.: Minimum determinant constraint for non-negative matrix factorization. Lecture Notes in Computer Science **Vol. 5441/2009**, 106–113 (2009)
17. Schmidt, M.N.: Linearly constrained bayesian matrix factorization for blind source separation. In: Y. Bengio, D. Schuurmans, J. Lafferty, C.K.I. Williams, A. Culotta (eds.) Advances in Neural Information Processing Systems, vol. 22, pp. 1624–1632 (2009)
18. Schmidt, M.N., Laurberg, H.: Nonnegative matrix factorization with gaussian process priors. Computational Intelligence and Neuroscience p. Article ID 361705 (2008)
19. Winter, M.E.: N-findr: an algorithm for fast autonomous spectral end-member determination in hyperspectral data. Proceedings of SPIE - The International Society for Optical Engineering **Vol.3753**, 266–275 (1999)