SUPER GAUSSIAN PRIORS FOR BLIND COLOR DECONVOLUTION OF HISTOLOGICAL IMAGES

Fernando Pérez-Bueno[1*] *Miguel Vega*[2] *Valery Naranjo*[3] *Rafael Molina*[1]
Aggelos K. Katsaggelos[4]

[1] Dpto. de Ciencias de la Computación e I.A. and [2] Dpto. de Lenguajes y Sistemas Informáticos, Universidad de Granada, Spain
[3]Dpto. de Comunicaciones, Universidad Politécnica de Valencia, Spain
[4]Dept. of Electrical and Computer Engineering, Northwestern University, Evanston, IL, USA

ABSTRACT

Color deconvolution aims at separating multi-stained images into single stained ones. In digital histopathological images, true stain color vectors vary between images and need to be estimated to obtain stain concentrations and separate stain bands. These band images can be used for image analysis purposes and, once normalized, utilized with other multi-stained images (from different laboratories and obtained using different scanners) for classification purposes. In this paper we propose the use of Super Gaussian (SG) priors for each stain concentration together with the similarity to a given reference matrix for the color vectors. Variational inference and an evidence lower bound are utilized to automatically estimate all the latent variables. The proposed methodology is tested on real images and compared to classical and state-of-the-art methods for histopathological blind image color deconvolution.

Index Terms— Blind Color Deconvolution, Histopathological Images, Variational Bayes, Super Gaussian

1. INTRODUCTION

Histopathological tissues are usually stained with a combination of stains that binds to specific proteins on the tissue. Hematoxylin and Eosin (H&E) is one of the most commonly used combination of stains. Hematoxylin stains cell nuclei while eosin stains cytoplasm and extracellular matrix components [5]. In digital brightfield microscopy, stained slides are then scanned to obtain high resolution Whole-Slide Images (WSI). Since the analysis of these images is very time consuming and requires a lot of effort, computer-aided diagnosis (CAD) systems have become a valuable ally for pathologists. These systems frequently make use of the information provided by the different stains separately [12]. The separation of the stains in a WSI is known as Color Deconvolution (CD) and aims at estimating each stain concentration at each pixel location. Usually, the color spectral properties of each stain are also unknown since they vary from image to image. Color variations have a wide range of origins: different scanners, stain manufactures, or staining procedures, among others. They create inter- and intra-laboratory differences. Blind Color Deconvolution (BCD) techniques estimate image specific stain color-vectors together with stain concentrations.

Several CD methods have been proposed (see [14] for a review). Ruifrok *et al.* [15] proposed one of pioneer works. Non-negative Matrix Factorization (NMF) [12, 18, 20], Singular Value Descomposition (SVD) [10, 11] and Independent Component Analysis (ICA) [1, 2, 17] have been applied to CD. Deep learning CD methods have also been recently proposed [8, 16, 21].

This paper contributes to the field with a general probabilistic framework to BCD. In [6, 7], a prior on the color-vectors, favouring similarity to some reference stain color-vectors, as well as a smoothness Simultaneous Autoregresive (SAR) prior model on each stain concentrations was used. As the SAR prior tends to oversmooth the edges of the image structures, the use of a Total Variation (TV) prior on each stain was proposed in [19]. Here we propose the use of Super Gaussian (SG) distributions as priors for BCD. SGs include distributions like l_p or log distributions which have proven to be very powerful prior models to induce sparsity in Bayesian Blind Deconvolution problems [3, 22]. They are always associated to energy functions with very interesting sparse properties. Furthermore, their inference procedure is easily carried out, as shown in section 3.

The rest of the paper is organized as follows: in section 2 the problem of BCD is mathematically formulated. Following the Bayesian modelling and inference, in section 3 we propose a fully Bayesian method for the estimation of the concentrations and the color-vector matrix. In section 4, the proposed method is evaluated and its performance is compared with other classical and state-of-the-art CD methods. Finally, section 5 concludes the work.

2. PROBLEM FORMULATION

Digital brightfield microscopes usually store a stained histological specimen's slide as an RGB color image of size $M \times N$, represented by the $MN \times 3$ matrix, $\mathbf{I}$. Each color plane is stacked into a $MN \times 1$ column vector $\mathbf{i}_c = (i_{1c}, \ldots, i_{MNc})^{\mathrm{T}}, c \in \{R, G, B\}$. Each value i_{ic} represents the transmitted light on color band $c \in \{R, G, B\}$ for the pixel i of the slide. However, for stain deconvolution is usual to work in the *Optical Density* (OD) space, where the Beer-Lamberth law, for a slide stained with n_s stains, establishes that

$$\mathbf{Y}^{\mathrm{T}} = \mathbf{M}\mathbf{C}^{\mathrm{T}} + \mathbf{N}^{\mathrm{T}}, \tag{1}$$

where $\mathbf{Y} \in \mathbb{R}^{MN \times 3}$ is the observed OD image with three OD channels, i.e., $\mathbf{Y} = [\mathbf{y}_R\ \mathbf{y}_G\ \mathbf{y}_B]$ and each channel $\mathbf{y}_c \in \mathbb{R}^{MN \times 1}$ is defined as $\mathbf{y}_c = -\log_{10}\left(\mathbf{i}_c/\mathbf{i}_c^0\right)$, where $\mathbf{i}_c^0$ denotes the incident light. The division operation and $\log_{10}(\cdot)$ function are computed elementwise. $\mathbf{M} \in \mathbb{R}^{3 \times n_s}$ is the color-vector matrix, $\mathbf{C} \in \mathbb{R}^{MN \times n_s}$ is the stain concentration matrix and $\mathbf{N} \in \mathbb{R}^{MN \times 3}$ is a random matrix with i.i.d. zero mean Gaussian components with variance β^{-1}.

*This work was sponsored in part by Ministerio de Ciencia e Innovación under Contract BES-2017-081584 and project DPI2016-77869-C2-2-R

Table 1. Some choices for the penalty function

Label	$\boldsymbol{\rho(s)}$	$\boldsymbol{\rho'(s)/\lvert s\rvert}$
$\ell_p\,, 0 < p \leq 1$	$\frac{1}{p}\lvert s\rvert^p$	$\lvert s\rvert^{p-2}$
log	$\log(\epsilon + \lvert s\rvert)$	$(\epsilon + \lvert s\rvert)^{-1}\lvert s\rvert^{-1}$

BCD techniques seek to estimate the stain concentration matrix.

$$\mathbf{C} = \begin{bmatrix} c_{11} & \dots & c_{1n_s} \\ \vdots & \ddots & \vdots \\ c_{MN1} & \dots & c_{MNn_s} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_{1,:}^{\mathrm{T}} \\ \vdots \\ \mathbf{c}_{MN,:}^{\mathrm{T}} \end{bmatrix} = \left[\mathbf{c}_1 \dots \mathbf{c}_{n_s}\right], \quad (2)$$

with the i-th row $\mathbf{c}_{i,:}^{\mathrm{T}} = (c_{i1}, \dots, c_{in_s})$, $i = 1, \dots, MN$, representing the contribution of each stain to the i-th $\mathbf{Y}$ pixel value, $\mathbf{y}_i$, and the s-th column $\mathbf{c}_s = (c_{1s}, \dots, c_{MNs})^{\mathrm{T}}$, $s \in \{1, \dots, n_s\}$, representing the concentrations of the s-th stain. BCD techniques also estimate the specific color of each stain, represented by the color-vector matrix $\mathbf{M} \in \mathbb{R}^{3\times n_s}$ where each column $\mathbf{m}_s$ in matrix $\mathbf{M}$ is a unit ℓ_2-norm stain color-vector containing the relative RGB color composition of the corresponding stain in the OD space.

In the following section we use Bayesian modeling and inference to estimate both $\mathbf{C}$ and $\mathbf{M}$.

3. BAYESIAN MODELLING AND INFERENCE

Following the degradation model in (1), we have

$$\begin{aligned} \mathrm{p}(\mathbf{Y}|\mathbf{C}, \mathbf{M}; \beta) &= \prod_{i=1}^{MN} \mathrm{p}(\mathbf{y}_{i,:}|\mathbf{M}, \mathbf{c}_{i,:}) \\ &= \prod_{i=1}^{MN} \mathcal{N}(\mathbf{y}_{i,:}|\mathbf{M}\mathbf{c}_{i,:}, \beta^{-1}\mathbf{I}_{3\times 3}) \end{aligned} \quad (3)$$

Bayesian methods start with a prior distribution on the unknowns. In this paper we adopt as priors SG distributions for the stain concentration in order to induce sparsity.

$$\begin{aligned} \mathrm{p}(\mathbf{C}; \boldsymbol{\alpha}) &= \prod_{\nu=1}^{L}\prod_{s=1}^{n_s} \mathrm{p}(\mathbf{c}_{\nu s}; \alpha_{\nu s}) \\ &\propto \prod_{\nu=1}^{L}\prod_{s=1}^{n_s}\prod_{i=1}^{MN} \exp\left[-\alpha_{\nu s}\rho(c_{\nu s}(i))\right], \end{aligned} \quad (4)$$

with $\alpha_{\nu s} > 0$. In (4) $\mathbf{c}_{\nu s} = \mathbf{D}_\nu \mathbf{c}_s$, where $\{\mathbf{D}_\nu\}_{\nu=1}^{L}$ is a set of L high-pass filters. For $\mathrm{p}(\mathbf{c}_{\nu s}; \alpha_{\nu s})$ in (4) to be SG $\rho(.)$ has to be symmetric around zero and the function $\rho(\sqrt{s})$ increasing and concave for $s \in (0, \infty)$. This condition is equivalent to $\rho'(s)/s$ being decreasing on $(0, \infty)$, and allows ρ to be represented as

$$\rho\left(c_{\nu s}(i)\right) = \inf_{\eta_{\nu s}(i)>0} \frac{1}{2}\,\eta_{\nu s}(i)\,{c_{\nu s}}^2(i) - \rho^*\left(\frac{1}{2}\eta_{\nu s}(i)\right) \quad (5)$$

$$\begin{aligned} \Rightarrow \rho\left(c_{\nu s}(i)\right) &\leq L\left(c_{\nu s}(i), \eta_{\nu s}(i)\right) \\ &= \frac{1}{2}\,\eta_{\nu s}(i)\,{c_{\nu s}}^2(i) - \rho^*\left(\frac{1}{2}\eta_{\nu s}(i)\right) \end{aligned} \quad (6)$$

where inf denotes the infimum, $\rho^*(\cdot)$ is the concave conjugate of $\rho(\cdot)$ and $\boldsymbol{\eta}_{\nu s} = \{\eta_{\nu s}(i)\}_{i=1}^{MN}$ are positive parameters. The relationship dual to (5) is given by [13]

$$\rho^*\left(\frac{1}{2}\eta_{\nu s}(i)\right) = \inf_{c_{\nu s}(i)} \frac{1}{2}\,\eta_{\nu s}(i)\,{c_{\nu s}}^2(i) - \rho\left(c_{\nu s}(i)\right). \quad (7)$$

Table 1 shows some penalty functions, corresponding to SG distributions (see [3] for additional energy functions associated to SG distributions). The use of different penalty functions will allow us to obtain different levels of sparsity on the concentration differences at neighbouring pixels. Notice that ideally, each stain fixes itself only and completely to certain proteins on the tissue, making the stain concentration differences at neighbouring pixels sparse [18].

For the unknown color-vector matrix $\mathbf{M} = [\mathbf{m}_1, \dots, \mathbf{m}_{n_s}]$ we incorporate as prior knowledge similarity to a reference color-vector matrix $\underline{\mathbf{M}} = [\underline{\mathbf{m}}_1, \dots, \underline{\mathbf{m}}_{n_s}]$ using

$$\mathrm{p}(\mathbf{M}; \boldsymbol{\gamma}) = \prod_{s=1}^{n_s} \mathrm{p}(\mathbf{m}_s; \gamma_s) \propto \prod_{s=1}^{n_s} \exp\left(-\frac{1}{2}\gamma_s \|\mathbf{m}_s - \underline{\mathbf{m}}_s\|^2\right), \quad (8)$$

where γ_s, $s = 1, \dots, n_s$, controls our confidence on the accuracy of $\underline{\mathbf{m}}_s$.

With all these ingredients, we define the joint probability distribution as

$$\begin{aligned} \mathrm{p}(\mathbf{Y}, \mathbf{C}, \mathbf{M}; \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}) =& \mathrm{p}(\mathbf{M}; \boldsymbol{\gamma})\mathrm{p}(\mathbf{Y}|\mathbf{C}, \mathbf{M}; \beta) \\ &\times \prod_{\nu=1}^{L}\prod_{s=1}^{n_s} \mathrm{p}(\mathbf{c}_{\nu s}; \alpha_{\nu s}). \end{aligned} \quad (9)$$

Following the Bayesian paradigm, inference will be based on the posterior distribution $\mathrm{p}(\Theta|\mathbf{Y}; \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ with $\Theta = \{\mathbf{C}, \mathbf{M}\}$ the set of all unknowns. In this paper we use the mean-field variational Bayesian model [4] to approximate $\mathrm{p}(\Theta|\mathbf{Y}; \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma})$ by the distribution $\mathrm{q}(\Theta)$ of the form $\mathrm{q}(\Theta) = \prod_{s=1}^{n_s} \mathrm{q}(\mathbf{m}_s) \prod_{\nu=1}^{L} \mathrm{q}(\mathbf{c}_{\nu s})$ that minimizes the Kullback-Leibler divergence [9] defined as

$$\mathbf{KL}\left(\mathrm{q}(\boldsymbol{\Theta}) \,||\, \mathrm{p}(\boldsymbol{\Theta}|\mathbf{Y})\right) = \int \mathrm{q}(\Theta) \log \frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta, \mathbf{Y})}\, \mathrm{d}\Theta + \log \mathrm{p}(\mathbf{Y}). \quad (10)$$

The Kullback-Leibler divergence is always non negative and equal to zero if and only if $\mathrm{q}(\Theta) = \mathrm{p}(\Theta|\mathbf{Y})$.

Even with this factorization, the SG prior for $\mathbf{C}_\nu$ hampers the evaluation of this divergence, but the quadratic bound for ρ in (6) allows us to bound the prior in (4) with a Gaussian form

$$\mathrm{p}\left(c_{\nu s}(i); \alpha_{\nu s}\right) \geq \exp[-\alpha_{\nu s} L(c_{\nu s}(i), \eta_{\nu s}(i))], \;\; \forall \eta_{\nu s}(i) > 0\,. \quad (11)$$

We then define $\mathcal{M}_\nu(\mathbf{C}, \boldsymbol{\eta}_\nu; \boldsymbol{\alpha}_\nu) = \prod_s \mathcal{M}_{\nu s}(\mathbf{c}_{\nu s}, \boldsymbol{\eta}_{\nu s}; \alpha_{\nu s})$ where

$$\mathcal{M}_{\nu s}(\mathbf{c}_{\nu s}, \boldsymbol{\eta}_{\nu s}|\alpha_{\nu s}) = \prod_{i=1}^{MN} \exp\left[-\alpha_{\nu s} L(c_{\nu s}(i), \eta_{\nu s}(i))\right] \quad (12)$$

and

$$\mathrm{F}(\Theta, \mathbf{U}, \mathbf{Y}) = \mathrm{p}(\mathbf{M}; \boldsymbol{\gamma})\mathrm{p}(\mathbf{Y}|\mathbf{C}, \mathbf{M}; \beta) \prod_{\nu=1}^{L} \mathcal{M}_\nu(\mathbf{C}, \boldsymbol{\eta}_\nu; \boldsymbol{\alpha}_\nu) \quad (13)$$

to obtain the inequality $\log \mathrm{p}(\Theta, \mathbf{Y}) \geq \log \mathrm{F}(\Theta, \mathbf{U}, \mathbf{Y})$.

Utilizing the lower bound $\mathrm{F}(\Theta, \mathbf{U}, \mathbf{Y})$ for the joint probability distribution in (10) we minimize $\mathbf{KL}(\mathrm{q}(\Theta) \,||\, \mathrm{F}(\Theta, \mathbf{U}\,, \mathbf{Y})$ instead of $\mathbf{KL}\left(\mathrm{q}(\Theta) \,||\, \mathrm{p}(\Theta|\mathbf{Y})\right)$.

As shown in [4], for each unknown $\theta \in \boldsymbol{\Theta}$, $\mathrm{q}(\theta)$ will have the form

$$\mathrm{q}(\theta) \propto \exp\left\langle \log \mathrm{p}(\mathbf{Y}, \mathbf{C}, \mathbf{M})\right\rangle_{\mathrm{q}(\boldsymbol{\Theta}\setminus\theta)}\,, \quad (14)$$

where $\boldsymbol{\Theta}\setminus\theta$ represents all the variables in $\boldsymbol{\Theta}$ except θ and $\langle\cdot\rangle_{\mathrm{q}(\Theta\setminus\theta)}$ denotes the expected value calculated using the distribution $\mathrm{q}(\boldsymbol{\Theta}\setminus\theta)$. When point estimates are required $\hat{\theta} = \langle\theta\rangle_{\mathrm{q}(\theta)}$ is used.

3.1. Concentration Update:

Defining the contribution of the s-th band

$$\mathbf{e}_{i,:}^{-s} = \mathbf{y}_{i,:} - \sum_{k\neq s} \langle c_{ik}\rangle \langle \mathbf{m}_k\rangle \quad \text{and}$$
$$z_i^{-s} = \langle \mathbf{m}_s\rangle^{\mathrm{T}} \mathbf{e}_{i,:}^{-s}, \quad i = 1,\ldots,MN, \tag{15}$$

we can easily show that

$$\langle \log \mathrm{p}(\mathbf{y},\mathbf{C},\mathbf{M};\beta,\boldsymbol{\alpha},\boldsymbol{\gamma})\rangle_{\mathrm{q}(\Theta\setminus\mathbf{c}_s)} =$$
$$-\frac{\beta}{2}\left(-2\mathbf{c}_s^{\mathrm{T}}\mathbf{z}^{-s} + \parallel \mathbf{c}_s\parallel^2 \left\langle \parallel \mathbf{m}_s\parallel^2\right\rangle\right)$$
$$-\sum_{\nu}\frac{\alpha_{\nu s}}{2}\mathbf{c}_s^{\mathrm{T}}\mathbf{D}_\nu^{\mathrm{T}}\mathrm{diag}(\boldsymbol{\eta}_{\nu s})\mathbf{D}_\nu\mathbf{c}_s + \text{const} \tag{16}$$

$\mathrm{q}(\mathbf{c}_s) = \mathcal{N}(\mathbf{c}_s|\langle\mathbf{c}_s\rangle, \boldsymbol{\Sigma}_{\mathbf{c}_s})$, where

$$\boldsymbol{\Sigma}_{\mathbf{c}_s}^{-1} = \beta\left\langle \parallel \mathbf{m}_s\parallel^2\right\rangle \mathbf{I}_{MN\times MN} + \sum_{\nu}\alpha_{\nu s}\mathbf{D}_\nu^{\mathrm{T}}\mathrm{diag}(\boldsymbol{\eta}_{\nu s})\mathbf{D}_\nu \tag{17}$$

$$\langle\mathbf{c}_s\rangle = \beta\boldsymbol{\Sigma}_{\mathbf{c}_s}\mathbf{z}^{-s}. \tag{18}$$

3.2. Color-Vector Update:

In a similar way, using (15), we calculate the distribution of $\mathbf{m}_s$,

$$\langle \log \mathrm{p}(\mathbf{y},\mathbf{C},\mathbf{M}|\boldsymbol{\beta},\boldsymbol{\alpha},\boldsymbol{\gamma})\rangle_{\mathrm{q}(\Theta\setminus\mathbf{m}_s)} =$$
$$-\frac{\beta}{2}\left(\parallel\mathbf{m}_s\parallel^2\sum_{i=1}^{MN}\langle c_{is}^2\rangle - 2\mathbf{m}_s^{\mathrm{T}}\sum_{i=1}^{MN}\langle c_{is}\rangle\mathbf{e}_{i,:}^{-s}\right)$$
$$-\frac{1}{2}\gamma_s\parallel\mathbf{m}_s-\underline{\mathbf{m}}_s\parallel^2 + \text{const}, \tag{19}$$

which produces

$$\mathrm{q}(\mathbf{m}_s) = \mathcal{N}(\mathbf{m}_s|\langle\mathbf{m}_s\rangle, \boldsymbol{\Sigma}_{\mathbf{m}_s}), \tag{20}$$

where

$$\boldsymbol{\Sigma}_{\mathbf{m}_s}^{-1} = \left(\sum_{\nu=1}^{L}\beta_\nu\sum_{i=1}^{MN}\langle c_{\nu is}^2\rangle + \gamma_s\right)\mathbf{I}_{3\times3},$$
$$\langle\mathbf{m}_s\rangle = \boldsymbol{\Sigma}_{\mathbf{m}_s}\left(\sum_{\nu=1}^{L}\beta_\nu\sum_{i=1}^{MN}\langle c_{\nu is}\rangle\mathbf{e}_{\nu i,:}^{-s} + \gamma_s\underline{\mathbf{m}}_s\right). \tag{21}$$

Notice that $\langle\mathbf{m}_s\rangle$ may not be a unitary vector even if $\underline{\mathbf{m}}_s$ is. We can always replace $\langle\mathbf{m}_s\rangle$ by $\langle\mathbf{m}_s\rangle / \parallel\langle\mathbf{m}_s\rangle\parallel$ and $\boldsymbol{\Sigma}_{\mathbf{m}_s}$ by $\boldsymbol{\Sigma}_{\mathbf{m}_s}/\parallel\langle\mathbf{m}_s\rangle\parallel^2$.

3.3. Variational Parameters Update:

To estimate the $\boldsymbol{\eta}$ matrix, we need to solve, for each $s\in\{1,\ldots,n_s\}$, $\nu\in\{1,\ldots,L\}$ and $i\in\{1,\ldots,MN\}$

$$\hat{\eta}_{\nu s}(i) = \arg\min_{\eta_{\nu_s}(i)}\langle L\left(c_{\nu s}(i),\eta_{\nu s}(i)\right)\rangle_{\mathrm{q}(\mathbf{c}_s)}$$
$$= \arg\min_{\eta_{\nu_s}(i)}\frac{1}{2}\eta_{\nu s}(i)\,u_{\nu s}^2(i) - \rho^*\left(\frac{1}{2}\eta_{\nu s}(i)\right) \tag{22}$$

where $u_{\nu s}(i) = \sqrt{\langle c_{\nu s}^2(i)\rangle}$.

Since

$$\rho^*\left(\frac{\hat{\eta}_{\nu s}(i)}{2}\right) = \min_x\frac{1}{2}\hat{\eta}_{\nu s}(i)x^2 - \rho(x) \tag{23}$$

whose minimum is achieved at $x = u_{\nu s}(i)$, we have, differentiating the right hand side of the above equation with respect to x,

$$\hat{\eta}_{\nu s}(i) = \rho'(u_{\nu s}(i))/u_{\nu s}(i). \tag{24}$$

3.4. Calculating the concentration covariance matrices:

The matrix $\boldsymbol{\Sigma}_{\mathbf{c}_s}$ must be explicitly calculated to find its trace and also $\hat{\eta}_{\nu s}(i)$. However, since its calculation is very intense, we propose the following approximation of the covariance matrix. We first approximate $\mathrm{diag}(\boldsymbol{\eta}_{\nu s})$ using

$$\mathrm{diag}(\boldsymbol{\eta}_{\nu s}) \approx z(\boldsymbol{\eta}_{\nu s})\mathbf{I}, \tag{25}$$

where $z(\boldsymbol{\eta}_{\nu s})$ is calculated as the mean of the values in the diagonal.

We then use the approximation

$$\boldsymbol{\Sigma}_{\mathbf{c}_s}^{-1} \approx \beta\left\langle\parallel\mathbf{m}_s\parallel^2\right\rangle\mathbf{I}_{MN\times MN} + \sum_{\nu}\alpha_{\nu s}z(\boldsymbol{\eta}_{\nu s})\mathbf{D}_\nu^{\mathrm{T}}\mathbf{D}_\nu = \mathbf{B}.$$

Finally we have

$$\left\langle c_{\nu s}^2(i)\right\rangle \approx \left(\langle c_{\nu s}(i)\rangle\right)^2 + \frac{1}{MN}\mathrm{tr}\left[\mathbf{B}^{-1}\mathbf{D}_\nu^{\mathrm{T}}\mathbf{D}_\nu\right].$$

Algorithm 1 Variational Bayesian SG Blind CD

Require: Observed image $\mathbf{I}$, reference (prior) color-vector matrix $\underline{\mathbf{M}}$, β, $\forall s$ γ_s and $\alpha_{\nu,s}$ $\forall\nu$.

Obtain the OD image $\mathbf{Y}$ from $\mathbf{I}$ and set $\langle\mathbf{m}_s\rangle^{(0)} = \underline{\mathbf{m}}_s$, $\boldsymbol{\Sigma}_{\mathbf{m}_s}^{(0)} = \mathbf{0}$, $\boldsymbol{\Sigma}_{\mathbf{c}_s}^{(0)} = \mathbf{0}$, $\langle\mathbf{c}_s\rangle^{(0)}$, $\forall s = 1,\ldots,n_s$, from the matrix $\mathbf{C}$ obtained as $\mathbf{C}^{\mathrm{T}} = \underline{\mathbf{M}}^{+}\mathbf{Y}^{\mathrm{T}}$, with $\underline{\mathbf{M}}^{+}$ the Moore-Penrose pseudo-inverse of $\underline{\mathbf{M}}$, and $n = 0$.

while convergence criterion is not met **do**

1. Set $n = n+1$.
2. Using $\langle\mathbf{c}_s\rangle^{(n-1)}$ and $\boldsymbol{\Sigma}_{\mathbf{c}_s}^{(n-1)}$ $\forall s$, update the new variational parameters $\hat{\boldsymbol{\eta}}_{\nu s}^{(n)}$ from (24) $\forall\nu$.
3. Using $\langle\mathbf{c}_s\rangle^{(n-1)}$, $\boldsymbol{\Sigma}_{\mathbf{c}_s}^{(n-1)}$ and $\langle\mathbf{m}_s\rangle^{(n-1)}$ update the color-vectors $\boldsymbol{\Sigma}_{\mathbf{m}_s}^{(n)}$ and $\langle\mathbf{m}_s\rangle^{(n)}$ from (21), $\forall s$.
4. Using $\langle\mathbf{m}_s\rangle^{(n)}$, $\boldsymbol{\Sigma}_{\mathbf{m}_s}^{(n)}$ and $\hat{\boldsymbol{\eta}}_{\nu s}^{(n)}$ $\forall\nu$ update the concentrations $\boldsymbol{\Sigma}_{\mathbf{c}_s}^{(n)}$ and $\langle\mathbf{c}_s\rangle^{(n)}$ from (17) and (18), $\forall s$.

end while

Output the color-vector $\hat{\mathbf{m}}_s = \langle\mathbf{m}_s\rangle^{(n)}$ and the concentrations $\hat{\mathbf{c}}_s = \langle\mathbf{c}_s\rangle^{(n)}$.

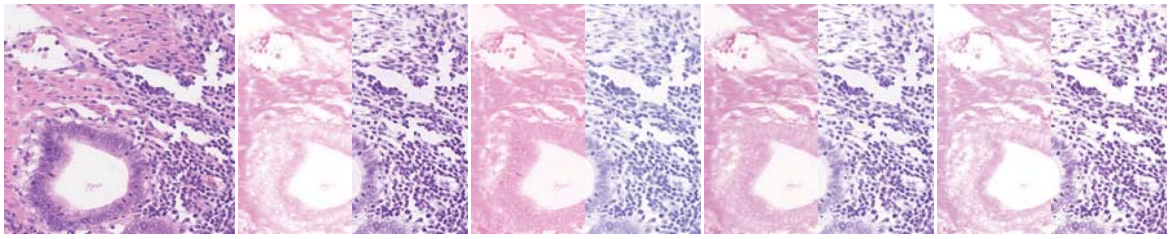
3.5. Proposed Algorithm

Based on the previous derivations, we propose the Variational Bayesian SG Blind Color Deconvolution in Algorithm 1. The linear equations problem in (18), used in step 3 of Alg. 1, has been solved using the Conjugate Gradient approach. Finally, from Alg. 1, an RGB image of each separated stain, $\hat{\mathbf{I}}_s^{\mathrm{sep}}$, can be obtained as $(\hat{\mathbf{I}}_s^{\mathrm{sep}})^{\mathrm{T}} = \exp_{10}(-\hat{\mathbf{m}}_s\hat{\mathbf{c}}_s^{\mathrm{T}})$.

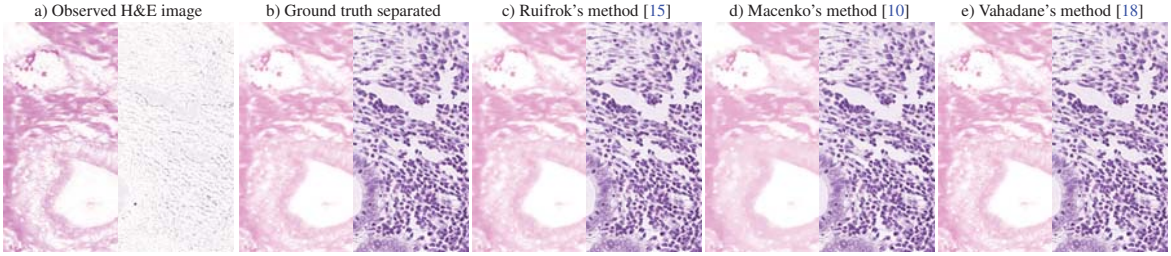
4. EXPERIMENTAL RESULTS

We compare the proposed approach with classical and state-of-the-art CD methods using the Warwick Stain Sçeparation Benchmark (WSSB) in [2]. WSSB includes 24 H&E stained images of breast, colon and lung tissues whose ground truth stain color-vector matrices, $\mathbf{M}_{GT}$, were manually selected based on biological structures. The ground truth concentrations were obtained using $\mathbf{C}_{GT}^{\mathrm{T}} = \mathbf{M}_{GT}^{+}\mathbf{Y}^{\mathrm{T}}$. An observed colon image is shown in Fig. 1a and its ground truth RGB separation is depicted in Fig. 1b.

The following parameter values $\beta = 6000$ and $\gamma_s = 10^{20}$ for $s = 1,2$ have been used for all images. Two penalty functions from



a) Observed H&E image b) Ground truth separated c) Ruifrok's method [15] d) Macenko's method [10] e) Vahadane's method [18]



f) Alsubaie's method [2] g) Hidalgo-Gavira's method [6] h) Vega's method [19] i) Proposed log prior j) Proposed ℓ_1 prior

Fig. 1. Colon observed H&E image from the WSSB dataset in [2], its ground truth separated E-only and H-only images and results for the competing and proposed methods. Eosin and hematoxylin separations are presented in the left and right hand side of each image, respectively.

Table 2. PSNR and SSIM for the different methods on the WSSB dataset [2].

Image	Stain	Ruifrok's method [15]		Macenko's method [10]		Vahadane's method [18]		Alsubaie's method [2]		Hidalgo-Gavira's method [6]		Vega's method [19]		Proposed log prior		Proposed ℓ_1 prior	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Colon	H	22.27	0.8141	23.91	0.8095	25.83	0.8851	21.11	0.7241	28.57	0.9542	28.62	0.9544	28.47	0.9511	**29.00**	**0.9638**
	E	20.70	0.7456	21.55	0.6365	26.29	0.8904	21.94	0.8540	27.58	0.9139	27.60	0.9161	27.55	0.9171	**28.38**	**0.9414**
Breast	H	15.27	0.6215	26.24	0.9552	25.46	0.9239	24.60	0.8068	28.81	0.9528	29.14	0.9560	29.44	0.9513	**30.50**	**0.9751**
	E	17.66	0.7644	23.62	0.9336	27.68	0.9550	25.92	0.9380	26.60	0.9464	26.76	0.9492	26.74	0.9444	**27.71**	**0.9645**
Lung	H	22.47	0.7987	19.52	0.7389	25.87	0.8912	20.62	0.5551	32.91	0.9763	33.10	0.9757	32.39	0.9681	**35.21**	**0.9898**
	E	22.05	0.7734	18.09	0.5088	25.53	0.8195	23.95	0.8939	30.77	0.9306	31.02	0.9353	30.61	0.9393	**33.07**	**0.9654**
Mean	H	20.00	0.7448	23.22	0.8345	25.72	0.9100	22.11	0.6953	30.10	0.9611	30.29	0.9621	30.10	0.9568	**31.57**	**0.9762**
	E	20.14	0.7611	21.08	0.6930	26.50	0.8883	23.94	0.8953	28.32	0.9303	28.46	0.9336	28.34	0.9347	**29.72**	**0.9571**

Table 1 have been used: log and ℓ_1. First order horizontal and vertical differences have been used as filters in (4), so $L = 2$. For $\nu = 1, 2$ the value $\alpha_{\nu,s} = 1.0$ for $s = 1, 2$ have been used with the log prior and the values $\alpha_{\nu,1} = 0.02$ and $\alpha_{\nu,2} = 0.01$ with the ℓ_1 prior. Algorithm 1 was run until the criterion $\| \langle \mathbf{c}_s \rangle^{(n)} - \langle \mathbf{c}_s \rangle^{(n-1)} \|^2 / \| \langle \mathbf{c}_s \rangle^{(n)} \|^2 < 10^{-3}$ was met by both stains, that is, $s = 1, 2$. Since different tissues may have different color characteristics, the reference color-vector matrix $\underline{\mathbf{M}}$ was obtained by selecting, by non-medical experts a single pixel containing mainly hematoxylin and another pixel containing mainly eosin from each type of breast, colon and lung tissues.

We compare our log and ℓ_1 models to the methods in [2,6,10,15, 18, 19]. For all the competing algorithms, parameters were selected following the recommendations in the original paper or the reference software freely available. The resulting H-only and E-only images were compared both visually and numerically by means of the Peak Signal to Noise Ratio (PSNR) and Structural Similarity (SSIM) metrics. Numerical results, presented in Table 2, show that the proposed ℓ_1 method produces higher PSNR and SSIM values than the competing models. The proposed log method produces similar results to the method in [19]. The separated H- and E-only images from the observed image in Fig. 1a are shown in Fig. 1c-j. The proposed method and the methods in [6, 18, 19] produce colors very similar to the ground truth separation in Fig. 1b although the new method and the one in [19] produce sharper images than the method in [6] and richer details than the method in [18].

The obtained results clearly show an advantage when using the ℓ_1 method, meaning that this prior captures the real behaviour of the dyes on the tissue. As we have already indicated, each stain fixes itself only and completely to certain proteins on the tissue, making the stain concentration differences at neighbouring pixels sparse [18]. However, the experiments show that the differences are not as sparse as expected. The ℓ_1 prior, with a lower kurtosis than log prior, allows to keep more non-zero values. This makes the ℓ_1 prior a good choice, as it induces sparsity in a softer way than the log prior.

5. CONCLUSIONS

In this work we have presented the use of Super Gaussian prior models for Blind Color Deconvolution. The use of the SG family induces sparsity on the differences of stain concentrations at neighbouring pixels. This is a desired quality during the staining process and a theoretical feature of the stained tissue. The variety of SG distributions available allows us to explore a range of possible sparse solutions within a common inference, keeping an easy procedure. In this work we explore two of the possible penalty functions, ℓ_1 and log. We find that the ℓ_1 prior, which has a more moderated peak, captures better the real distribution of the stains present in the images. The softer sparsity induction leads us to a blind color deconvolution method that outperforms all the methods compared with. Future work using this promising methodology includes model parameter estimation, and the use of this new blind CD method as preprocessing for WSI classification.

6. REFERENCES

[1] Alsubaie, N., Raza, S.E.A., Rajpoot, N.: Stain deconvolution of histology images via independent component analysis in the wavelet domain. In: ISBI. pp. 803–806 (2016)

[2] Alsubaie, N., Trahearn, N., et al.: Stain deconvolution using statistical analysis of multi-resolution stain colour representation. PLOS ONE **12**, e0169875 (2017)

[3] Babacan, S.D., Molina, R., Do, M.N., Katsaggelos, A.K.: Blind deconvolution with general sparse image priors. In: ECCV (2012)

[4] Bishop, C.: Pattern Recognition and Machine Learning, pp. 454–455. Springer (2006)

[5] Fischer, A.H., Jacobson, K.A., Rose, J., Zeller, R.: Hematoxylin and Eosin Staining of Tissue and Cell Sections. Cold Spring Harbor Protocols (2008)

[6] Hidalgo-Gavira, N., Mateos, J., Vega, M., Molina, R., Katsaggelos, A.K.: Variational bayesian blind color deconvolution of histopathological images. IEEE Transactions on Image Processing **29**(1), 2026–2036 (2020)

[7] Hidalgo-Gavira, N., Mateos, J., et al.: Blind color deconvolution of histopathological images using a variational Bayesian approach. In: ICIP. pp. 983–987 (2018)

[8] Janowczyk, A., Basavanhally, A., Madabhushi, A.: Stain normalization using sparse autoencoders (StaNoSA): Application to digital pathology. Computerized Medical Imaging and Graphics **57**, 50–61 (2017)

[9] Kullback, S.: Information Theory and Statistics. Dover Pub. (1959)

[10] Macenko, M., Niethammer, M., et al.: A method for normalizing histology slides for quantitative analysis. In: ISBI. pp. 1107–1110 (2009)

[11] McCann, M.T., Majumdar, J., et al.: Algorithm and benchmark dataset for stain separation in histology images. In: ICIP. pp. 3953–3957 (2014)

[12] Rabinovich, A., Agarwal, S., et al.: Unsupervised color decomposition of histologically stained tissue samples. In: NIPS. pp. 667–674 (2004)

[13] Rockafellar, R.: Convex analysis. Princeton University Press (1996)

[14] Roy, S., Jain, A.K., et al.: A study about color normalization methods for histopathology images. Micron **114**, 42–61 (2018)

[15] Ruifrok, A.C., Johnston, D.A.: Quantification of histochemical staining by color deconvolution. Analytical and quantitative cytology and histology **23**, 291–299 (2001)

[16] Shaban, M.T., Baur, C., Navab, N., Albarqouni, S.: Staingan: Stain style transfer for digital histological images. In: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). pp. 953–956. IEEE (2019)

[17] Trahearn, N., Snead, D., et al.: Multi-class stain separation using independent component analysis. In: Medical Imaging 2015: Digital Pathology. p. 94200J (2015)

[18] Vahadane, A., Peng, T., et al.: Structure-preserving color normalization and sparse stain separation for histological images. IEEE Trans. on Medical Imaging **35**, 1962–1971 (2016)

[19] Vega, M., Mateos, J., Molina, R., Katsaggelos, A.K.: Variational bayes color deconvolution with a total variation prior. In: 2019 27th European Signal Processing Conference (EUSIPCO). pp. 1–5 (Sep 2019)

[20] Xu, J., Xiang, L., et al.: Sparse non-negative matrix factorization (SNMF) based color unmixing for breast histopathological image analysis. Computerized Medical Imaging and Graphics **46**, 20–29 (2015)

[21] Zanjani, F.G., Zinger, S., Bejnordi, B.E., Laak, J.A.W.M.v.d., With, P.H.N.d.: Stain normalization of histopathology images using generative adversarial networks. In: 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018). pp. 573–577 (Apr 2018)

[22] Zhou, X., Vega, M., Zhou, F., Molina, R., Katsaggelos, A.K.: Fast bayesian blind deconvolution with huber super gaussian priors. Digital Signal Processing **60**, 122–133 (2017)