VARIATIONAL BAYESIAN COMPRESSIVE BLIND IMAGE DECONVOLUTION

Bruno Amizic[1], *Leonidas Spinoulas*[1], *Rafael Molina*[2*], *Aggelos K. Katsaggelos*[1†]

[1] Dept. of Electrical Engineering and Comp. Sc., Northwestern University, Evanston, IL 60208, USA
[2] Dept. de Ciencias de la Computación e I.A., Universidad de Granada, 18071 Granada, Spain
Email: amizic@northwestern.edu, leonisp@u.northwestern.edu, rms@decsai.ugr.es, aggk@eecs.northwestern.edu

ABSTRACT

We propose a novel variational Bayesian framework to perform simultaneous compressive sensing (CS) image reconstruction and blind deconvolution (BID) as well as estimate all modeling parameters. Furthermore, we show that the proposed framework generalizes the alternating direction method of multipliers which is often utilized to transform a constrained optimization problem into an unconstrained one through the use of the augmented Lagrangian. The proposed framework can be easily adapted to other signal processing applications or particular image and blur priors within the proposed context. In this work, as an example, we employ the following priors to illustrate the significance of the proposed approach: (1) a non-convex l_p quasi-norm based prior for the image, (2) a simultaneous auto-regressive prior for the blur, and (3) an l_1 norm based prior for the transformed coefficients. Experimental results using synthetic images demonstrate the advantages of the proposed algorithm over existing approaches.

Index Terms— Inverse methods, compressive sensing, blind image deconvolution, parameter estimation, Bayesian methods.

1. INTRODUCTION

Recently, the compressive BID problem has been introduced [1–3]. The objective of this inverse problem is to estimate both the image and blur from compressed measurements. The acquisition model for the compressed lens-based measurements in matrix-vector form is given by

$$\mathbf{y} = \mathbf{\Phi}\mathbf{H}\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{\Phi}$ corresponds to the $M \times N$ CS measurement matrix ($M \ll N$), the $N \times N$ matrix $\mathbf{H}$ denotes the convolution matrix resulting from the point spread function (PSF) of the lens, limited aperture dimensions, lack of focus, atmospheric turbulences or combinations of the above, the $N \times 1$ vector $\mathbf{x}$ denotes the unknown image and the $M \times 1$ vector $\mathbf{n}$ represents the measurement noise. The authors in [2] derived the lower bound for the number of measurements needed to accurately reconstruct a sparse vector $\mathbf{x}$ and proposed a reconstruction algorithm that recovers simultaneously $\mathbf{x}$ and $\mathbf{h}$ for several real-world acquisition systems (e.g., neuronal pulse streams and astronomical imaging). One of the disadvantages of the method in [2] is its inability to recover the unknown vectors when pulses overlap.

In our earlier work, [3], we showed experimentally that a blurred image (i.e., $\mathbf{Hx}$) is compressible in various scenarios and proposed a constrained optimization framework for compressive BID. Compressible signals (see [4, 5]) can be well approximated by a linear combination of K ($K \ll N$) basis vectors from the column space of a transformation matrix $\mathbf{W}$ (i.e., $\mathbf{Wa} \approx \mathbf{Hx}$), where the $N \times 1$ vector $\mathbf{a}$ corresponds to the sparse transformed coefficients of the blurred signal.

One deficiency of our earlier work, presented in [3], is the assumption that the model parameters yielding best restoration results are known *a priori*. However, in general, optimal parameters depend on the observation noise, measurement matrix, compressive ratio, blur, and the particular image of interest, among other factors. In this paper we extend our framework from [3] by introducing a novel variational Bayesian CS BID algorithm for which all model parameters and unknown vectors are simultaneously inferred. Furthermore, we show how to generalize the alternating direction method of multipliers (ADMM, see [6] for details) to include the estimation of the unknown regularization parameters through the use of the proposed framework.

This paper is organized as follows. We introduce a hierarchical Bayesian framework and explain the basis of the proposed approach in Section 2. The proposed variational Bayesian CS BID algorithm is presented in Section 3 while simulation results that support our approach are shown in Section 4. Finally, conclusions are drawn in Section 5.

2. HIERARCHICAL BAYESIAN MODELING

The measurement noise is modeled as a zero mean white Gaussian random vector. Therefore, the observation model is

*This work was supported in part by the Spanish Ministry of Economy and Competitiveness under project TIN2010-15137, by the European Regional Development Fund (FEDER) and by the CEI BioTic at the Universidad de Granada.

†This work was supported in part by the Department of Energy under contract DE-NA0000457.

defined as,

$$\mathrm{p}(\mathbf{y}|\beta,\mathbf{a}) \propto \beta^{\frac{M}{2}} \exp\left[-\frac{\beta}{2}\|\mathbf{y}-\mathbf{\Phi}\mathbf{W}\mathbf{a}\|^2\right], \quad (2)$$

where β is the precision of the multivariate Gaussian distribution.

In this work, similarly to [7], we utilize a variant of the generalized Gaussian distribution (see [8] for more details) for the image prior since the derivatives of blurry photographs are expected to be sparse. More specifically,

$$\mathrm{p}(\mathbf{x}|\alpha) \propto \alpha^{\lambda_1 \frac{N}{p}} \exp\left[-\alpha \sum_{d\in \mathrm{D}} 2^{1-o(d)} \sum_{i=1}^{N} |\Delta_i^d(\mathbf{x})|^p\right], \quad (3)$$

where λ_1 is a positive real number, $\mathrm{D} = \{h, v, hh, vv, hv\}$, $0 < p < 1$ and $o(d) \in \{1, 2\}$ denotes the order of the difference operator $\Delta_i^d(\mathbf{x})$. $\Delta_i^h(\mathbf{x})$ and $\Delta_i^v(\mathbf{x})$ correspond, respectively, to the horizontal and vertical first order differences, at pixel i. That is, $\Delta_i^h(\mathbf{x}) = x_i - x_{l(i)}$ and $\Delta_i^v(\mathbf{x}) = x_i - x_{a(i)}$, where $l(i)$ and $a(i)$ denote the nearest neighbors of i, to the left and above, respectively. The operators $\Delta_i^{hh}(\mathbf{x})$, $\Delta_i^{vv}(\mathbf{x})$, $\Delta_i^{hv}(\mathbf{x})$ correspond, respectively, to horizontal, vertical and horizontal-vertical second order differences, at pixel i.

The blur is modeled by a simultaneous-autoregression,

$$\mathrm{p}(\mathbf{h}|\gamma) \propto \gamma^{\frac{N}{2}} \exp\left[-\frac{\gamma}{2}\|\mathbf{C}\mathbf{h}\|^2\right], \quad (4)$$

where $\mathbf{C}$ denotes the Laplacian operator. The transformed coefficients $\mathbf{a}$ are modeled by a multivariate Laplace prior

$$\mathrm{p}(\mathbf{a}|\tau) = \left(\frac{\tau}{2}\right)^N \exp\left[-\tau\|\mathbf{a}\|_1\right]. \quad (5)$$

Finally, for $\omega \in \{\alpha, \beta, \gamma, \tau\}$, we use flat improper hyperpriors such that,

$$\mathrm{p}(\omega) \propto \text{const.} \quad (6)$$

The joint distribution of all unknowns can be found by combining (2), (3), (4), (5), and (6) as,

$$\mathrm{p}(\Theta, \mathbf{y}) \propto \mathrm{p}(\mathbf{y}|\beta,\mathbf{a})\mathrm{p}(\mathbf{x}|\alpha)\mathrm{p}(\mathbf{h}|\gamma)\mathrm{p}(\mathbf{a}|\tau), \quad (7)$$

where $\Theta = \{\alpha, \beta, \gamma, \tau, \mathbf{a}, \mathbf{x}, \mathbf{h}\}$. Note that the model in (7) does not include the dependency between the unknown vectors (i.e., $\mathbf{a}$, $\mathbf{x}$, and $\mathbf{h}$) shown in our previous work (see [3]). In [3], we solve the constrained optimization problem

$$\begin{aligned} &\underset{\mathbf{x},\mathbf{h},\mathbf{a}}{\text{minimize}} \quad \frac{\beta}{2}\|\mathbf{y}-\mathbf{\Phi}\mathbf{W}\mathbf{a}\|^2 + \tau\|\mathbf{a}\|_1 + \alpha R_1(\mathbf{x}) + \frac{\gamma}{2} R_2(\mathbf{h}) \\ &\text{subject to} \quad \mathbf{H}\mathbf{x} = \mathbf{W}\mathbf{a}, \end{aligned} \quad (8)$$

where $R_1(\mathbf{x})$ and $R_2(\mathbf{h})$ are the regularization functionals associated with the proposed image and blur prior, respectively. We can solve (8) by utilizing the ADMM (see [6]) which iteratively introduces scaled Lagrangian multipliers $\mathbf{u}$ at each iteration k, such that, $\mathbf{u}^{k+1} = \mathbf{u}^k + (\mathbf{W}\mathbf{a}^{k+1} - \mathbf{H}^{k+1}\mathbf{x}^{k+1})$.

In order to include the modeling of the Lagrangian multipliers in the proposed Bayesian framework, let us define the pseudo-observation model

$$\mathrm{p}(\mathbf{u}|\mathbf{a},\mathbf{x},\mathbf{h}) \propto \eta^{\frac{N}{2}} \exp\left[-\frac{\eta}{2}\|\mathbf{W}\mathbf{a}-\mathbf{H}\mathbf{x}+\mathbf{u}\|^2\right], \quad (9)$$

where η is a known positive real number. Now, the joint distribution of all unknowns can finally be found by combining (2), (3), (4), (5), (6) and (9) as,

$$\mathrm{p}(\Theta, \mathbf{y}, \mathbf{u}) \propto \mathrm{p}(\mathbf{y}|\beta,\mathbf{a})\mathrm{p}(\mathbf{u}|\mathbf{a},\mathbf{x},\mathbf{h})\mathrm{p}(\mathbf{x}|\alpha)\mathrm{p}(\mathbf{h}|\gamma)\mathrm{p}(\mathbf{a}|\tau). \quad (10)$$

3. BAYESIAN INFERENCE AND VARIATIONAL APPROXIMATION

Bayesian inference on the unknown components of the compressive BID problem is based on the estimation of the unknown posterior distribution $\mathrm{p}(\Theta|\mathbf{y},\mathbf{u})$, given by

$$\mathrm{p}(\Theta|\mathbf{y},\mathbf{u}) = \frac{\mathrm{p}(\Theta,\mathbf{y},\mathbf{u})}{\mathrm{p}(\mathbf{y},\mathbf{u})}. \quad (11)$$

Since the direct computation of $\mathrm{p}(\Theta|\mathbf{y},\mathbf{u})$ is intractable we resort to variational methods to approximate it by a tractable distribution of the form

$$\mathrm{q}(\Theta) = \mathrm{q}(\alpha,\beta,\gamma,\tau)\mathrm{q}(\mathbf{a})\mathrm{q}(\mathbf{x})\mathrm{q}(\mathbf{h}). \quad (12)$$

The variational criterion used to find $\mathrm{q}(\Theta)$ is the minimization of the Kullback-Leibler (KL) divergence [9], given by

$$\begin{aligned} &C_{KL}\left(\mathrm{q}(\Theta) \parallel \mathrm{p}(\Theta|\mathbf{y},\mathbf{u})\right) = \int \mathrm{q}(\Theta)\log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta|\mathbf{y},\mathbf{u})}\right) d\Theta \\ &= \int \mathrm{q}(\Theta)\log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{p}(\Theta,\mathbf{y},\mathbf{u})}\right) d\Theta + \text{const} \\ &= \mathcal{M}\left(\mathrm{q}(\Theta)\right) + \text{const}, \end{aligned} \quad (13)$$

which is always non negative and equal to zero only when the distributions $\mathrm{q}(\Theta)$ and $\mathrm{p}(\Theta|\mathbf{y},\mathbf{u})$ coincide.

Unfortunately, the KL divergence, defined in (13), cannot be evaluated due to the form of the image prior. To overcome this difficulty, we employ the majorization-minimization approach in [10,11] to find an upper bound of the KL divergence which allows for analytical Bayesian inference. We start by bounding the non-convex image prior $\mathrm{p}(\mathbf{x}|\alpha)$ (see [8] for detailed derivation) as,

$$\mathrm{p}(\mathbf{x}|\alpha) \geq \text{const} \cdot \mathrm{M}(\alpha,\mathbf{x},\mathbf{V}), \quad (14)$$

where $\mathbf{V}$ is a matrix with elements $v_{d,i} > 0$ for $d \in \{h, v, hh, vv, hv\}$, $i = 1, \ldots, N$ and

$$\mathrm{M}(\alpha,\mathbf{x},\mathbf{V}) = \alpha^{\lambda_1 \frac{N}{p}} \exp\left[-\frac{\alpha p}{2} \sum_{d\in\mathrm{D}} 2^{1-o(d)} \sum_i \frac{[\Delta_i^d(\mathbf{x})]^2 + \frac{2-p}{p} v_{d,i}}{v_{d,i}^{1-\frac{p}{2}}}\right]. \quad (15)$$

The inequality in (14) leads to the following lower bound for the joint probability distribution,

$$\mathrm{p}(\Theta,\mathbf{y},\mathbf{u}) \geq \mathrm{F}(\Theta,\mathbf{V},\mathbf{y},\mathbf{u}), \tag{16}$$

where

$$\begin{aligned}\mathrm{F}(\Theta,\mathbf{V},\mathbf{y},\mathbf{u}) = \mathrm{const}\cdot \mathrm{p}(\mathbf{y}|\beta,\mathbf{a})\mathrm{p}(\mathbf{u}|\mathbf{a},\mathbf{x},\mathbf{h})\\ \mathrm{M}(\alpha,\mathbf{x},\mathbf{V})\mathrm{p}(\mathbf{h}|\gamma)\mathrm{p}(\mathbf{a}|\tau).\end{aligned} \tag{17}$$

By defining the KL divergence between the distribution $\mathrm{q}(\Theta)$ and the bound $\mathrm{F}(\Theta,\mathbf{V},\mathbf{y},\mathbf{u})$, that is,

$$\tilde{\mathcal{M}}(\mathrm{q}(\Theta),\mathbf{V}) = \int \mathrm{q}(\Theta)\log\left(\frac{\mathrm{q}(\Theta)}{\mathrm{F}(\Theta,\mathbf{V},\mathbf{y},\mathbf{u})}\right)d\Theta\,, \tag{18}$$

and by utilizing inequality (16) we obtain

$$\mathcal{M}(\mathrm{q}(\Theta)) \leq \min_{\mathbf{V}} \tilde{\mathcal{M}}(\mathrm{q}(\Theta),\mathbf{V}), \tag{19}$$

where the upper bound in (19), with the addition of the unknown variational matrix $\mathbf{V}$, is mathematically tractable and therefore can be used to calculate the posterior distribution $\mathrm{q}(\Theta)$ by minimizing (18).

The differentiation of (18) with respect to $\mathrm{q}(\theta)$, with $\theta\in\Theta$, results to the posterior

$$\hat{\mathrm{q}}(\theta) \propto \exp\left[\langle \log \mathrm{F}(\Theta,\mathbf{V},\mathbf{y},\mathbf{u})\rangle_{\mathrm{q}(\Theta_\theta)}\right], \tag{20}$$

where the set Θ_θ denotes the result of the set difference $\Theta\backslash\{\theta\}$ and the operator $\langle\cdot\rangle_{\mathrm{q}(\Theta_\theta)}$ denotes the expected value with respect to the distribution $\mathrm{q}(\Theta_\theta)$. For convenience, from now on, wherever is obvious, the distribution used for the calculation of the expected value will be dropped from the notation (e.g., $\langle\mathbf{a}\rangle$ denotes $\langle\mathbf{a}\rangle_{\mathrm{q}(\mathbf{a})}$).

Next, we assume that the distributions $\mathrm{q}(\mathbf{x})$ and $\mathrm{q}(\mathbf{a})$ are degenerate, that is, they take one value with probability one and the rest of the values with probability zero, and that the distributions $\mathrm{q}(\alpha,\beta,\gamma,\tau)$ and $\mathrm{q}(\mathbf{h})$ are not constrained to be degenerate. In the experimental section we show that the degenerate assumption for the image $\mathbf{x}$ and the transformed coefficients $\mathbf{a}$ is sufficient to outperform current state-of-the-art methods. Now, let us propose the compressive BID algorithm based on the variational Bayesian framework.

Algorithm. *Given η, $\lambda_1 = 0.5$, $p = 0.8$, $\mathbf{x}^{1,1} = \mathbf{W}\mathbf{a}^1$, $\mathbf{u}^1 = 0*\mathbf{a}^1$, and initial estimates of the mean and covariance for the blur posterior $\langle\mathbf{h}\rangle^{1,1}$ and $\boldsymbol{\Sigma}_{\mathbf{h}}^{1,1} = 0*\mathbf{I}$, respectively.*

For $k = 1,2,\ldots$ until a stopping criterion is met:

1. *For $l = 1,2,\ldots$ until a stopping criterion is met:*

1.a) Calculate

$$\langle\alpha\rangle^{k,l} = \frac{\lambda_1\frac{N}{p}+1}{\sum_{d\in\mathrm{D}} 2^{1-o(d)}\sum_i |\Delta_i^d(\mathbf{x}^{k,l})|^p}, \tag{21}$$

$$\langle\gamma\rangle^{k,l} = \frac{N+2}{\langle\|\,\mathbf{C}\mathbf{h}\,\|^2\rangle_{\mathrm{q}^{k,l}(\mathbf{h})}}. \tag{22}$$

1.b) For each $d\in\{h,v,hh,vv,hv\}$ calculate

$$v_{d,i}^{k,l} = [\Delta_i^d(\mathbf{x}^{k,l})]^2. \tag{23}$$

1.c) Calculate

$$\begin{aligned}\mathbf{x}^{k,l} &= \arg\min_{\mathbf{x}}\Big\{\frac{\eta}{2}\langle\|\mathbf{W}\mathbf{a}-\mathbf{H}\mathbf{x}+\mathbf{u}^k\|^2\rangle_{\mathrm{q}^{k,l}(\mathbf{h})\mathrm{q}^k(\mathbf{a})}+\\ &\frac{\langle\alpha\rangle^{k,l}p}{2}\sum_{d\in\mathrm{D}}2^{1-o(d)}\sum_i\frac{[\Delta_i^d(\mathbf{x})]^2+\frac{2-p}{p}v_{d,i}}{v_{d,i}^{1-\frac{p}{2}}}\Big\} = \\ &= \Big[\eta(\langle\mathbf{H}\rangle^{k,l})^t(\langle\mathbf{H}\rangle^{k,l})+\eta N\boldsymbol{\Sigma}_h^{k,l}\\ &\langle\alpha\rangle^{k,l}p\sum_d 2^{1-o(d)}(\boldsymbol{\Delta}^d)^t\mathrm{B}_d^{k,l}(\boldsymbol{\Delta}^d)\Big]^{-1}\times\\ &\eta(\langle\mathbf{H}\rangle^{k,l})^t(\mathbf{W}\langle\mathbf{a}\rangle^k+\mathbf{u}^k),\end{aligned} \tag{24}$$

where $\boldsymbol{\Delta}^d$ is the convolution matrix of the difference operator $\Delta_i^d(\cdot)$ and $\mathrm{B}_d^{k,l}$ is a diagonal matrix with entries $\mathrm{B}_d^{k,l}(i,i) = (v_{d,i}^{k,l})^{\frac{p}{2}-1}$.

1.d) Calculate

$$\begin{aligned}\langle\mathbf{h}\rangle^{k,l+1} &= \arg\min_{\mathbf{h}}\Big\{\frac{\eta}{2}\langle\|\mathbf{W}\mathbf{a}-\mathbf{H}\mathbf{x}+\mathbf{u}^k\|^2\rangle_{\mathrm{q}^{k,l}(\mathbf{x})\mathrm{q}^k(\mathbf{a})}+\\ &\frac{\langle\gamma\rangle^{k,l}}{2}\|\mathbf{C}\mathbf{h}\|^2\Big\} = \eta\boldsymbol{\Sigma}_{\mathbf{h}}^{k,l+1}(\mathbf{X}^{k,l})^t(\mathbf{W}\langle\mathbf{a}\rangle^k+\mathbf{u}^k),\end{aligned} \tag{25}$$

where $\boldsymbol{\Sigma}_{\mathbf{h}}^{k,l+1} = \left[\eta(\mathbf{X}^{k,l})^t(\mathbf{X}^{k,l})+\langle\gamma\rangle^{k,l}\mathbf{C}^t\mathbf{C}\right]^{-1}$ and $\mathbf{X}^{k,l}$ is the convolution matrix of the image $\mathbf{x}^{k,l}$.

2. *Set $\mathbf{x}^{k+1,1} = \mathbf{x}^k = \mathbf{x}^{k,l}$, $\langle\mathbf{h}\rangle^{k+1,1} = \langle\mathbf{h}\rangle^k = \langle\mathbf{h}\rangle^{k,l+1}$, $\boldsymbol{\Sigma}_{\mathbf{h}}^{k+1,1} = \boldsymbol{\Sigma}_{\mathbf{h}}^k = \boldsymbol{\Sigma}_{\mathbf{h}}^{k,l+1}$.*

$$\mathrm{q}^{k+1,1}(\mathbf{x}) = \mathrm{q}^k(\mathbf{x}) = \begin{cases}1, & for\quad \mathbf{x}=\mathbf{x}^k\\ 0, & otherwise\end{cases}. \tag{26}$$

$$\mathrm{q}^{k+1,1}(\mathbf{h}) = \mathrm{q}^k(\mathbf{h}) = \mathcal{N}\left(\langle\mathbf{h}\rangle^k,\boldsymbol{\Sigma}_{\mathbf{h}}^k\right), \tag{27}$$

where $\mathcal{N}(\cdot,\cdot)$ is a Gaussian multivariate distribution.

3. *Calculate*

$$\langle\beta\rangle^k = \frac{M+2}{\langle\|\,\mathbf{y}-\boldsymbol{\Phi}\mathbf{W}\mathbf{a}\,\|^2\rangle^k}, \tag{28}$$

$$\langle\tau\rangle^k = \frac{N+1}{\langle\|\mathbf{a}\|_1\rangle^k}. \tag{29}$$

4. *Find*

$$\begin{aligned}\mathbf{a}^{k+1} &= \arg\min_{\mathbf{a}}\Big\{\frac{\langle\beta\rangle^k}{2}\,\|\,\mathbf{y}-\boldsymbol{\Phi}\mathbf{W}\mathbf{a}\,\|^2+\\ &+\frac{\eta}{2}\langle\|\,\mathbf{W}\mathbf{a}-\mathbf{H}\mathbf{x}+\mathbf{u}^k\,\|^2\rangle_{\mathrm{q}^k(\mathbf{x})\mathrm{q}^k(\mathbf{h})}+\langle\tau\rangle^k\|\mathbf{a}\|_1\Big\}.\end{aligned} \tag{30}$$

5. *Set*

$$\mathrm{q}^{k+1}(\mathbf{a}) = \begin{cases}1, & for\quad \mathbf{a}=\mathbf{a}^{k+1}\\ 0, & otherwise\end{cases}. \tag{31}$$

6. *Calculate*

$$\mathbf{u}^{k+1} = \mathbf{u}^k + (\mathbf{W}\mathbf{a}^{k+1}-\langle\mathbf{H}\rangle^k\mathbf{x}^k). \tag{32}$$

Note that, since (30) can be re-written as,

$$\mathbf{a}^{k+1} = \underset{\mathbf{a}}{\operatorname{argmin}} \parallel \mathbf{y}' - \mathbf{\Phi}'\mathbf{W}\mathbf{a} \parallel^2 + \langle\tau\rangle^k \|\mathbf{a}\|_1, \qquad (33)$$

where $\mathbf{y}' = \left[\sqrt{\frac{\langle\beta\rangle^k}{2}}\mathbf{y}^t \quad \left[\sqrt{\frac{\eta}{2}}(\langle\mathbf{H}\rangle^k\langle\mathbf{x}\rangle^k - \mathbf{u}^k)\right]^t\right]^t$ and $\mathbf{\Phi}' = \left[\sqrt{\frac{\langle\beta\rangle^k}{2}}\mathbf{\Phi}^t \quad \sqrt{\frac{\eta}{2}}\mathbf{I}\right]^t$, then when solving for the sparse vector $\mathbf{a}$, existing CS reconstruction algorithms (e.g., [12]) can be employed. Finally, as mentioned earlier, different image and blur regularizers can be used in our framework. This makes the proposed approach a versatile model for solving compressive BID problems.

Let us point out another interesting fact about the proposed algorithm. For simplicity's sake, let us assume that the hyperparameters (i.e., α, β, γ, and τ) are known and that the posterior distribution $\mathrm{q}(\mathbf{h})$ takes the degenerate form. Now, by iteratively updating the pseudo-observation $\mathbf{u}$ as shown in (32), our algorithm coincides with the ADMM (see [6]) which can be utilized to solve (8). Therefore, the proposed algorithm generalizes the ADMM by estimating the parameters of their respective priors or regularization terms.

4. EXPERIMENTAL RESULTS

In this section we conduct a series of synthetic experiments to support our CS BID approach. For comparison purposes we use a set of state-of-the art algorithms, namely *l1-ls* [12], *GPSR* [13] and *NESTA* [14] (for non-blind experiments) as well as our earlier approach [3] (denoted as *CSBD-S*).

We use the "Cameraman", "Lena", and "Shepp-Logan" images of size 256×256 pixels. All images are normalized, degraded with a Gaussian PSF and Gaussian noise is added to the CS measurements. As a performance metric, we use the peak signal to noise ratio (PSNR) which is defined as $\text{PSNR} = 10\log_{10}\frac{NL^2}{\|\mathbf{x}-\hat{\mathbf{x}}\|^2}$, where $\mathbf{x}$ and $\hat{\mathbf{x}}$ are the original and estimated images, respectively, and the constant L represents the maximum possible intensity value in image $\mathbf{x}$. Furthermore, for all experiments, we use the 3-level Haar Wavelet transform as our sparsifying basis $\mathbf{W}$.

We perform two series of experiments for each image: non-blind reconstructions/restorations when the blur $\mathbf{h}$ is assumed to be known, and blind reconstructions/restorations when the blur $\mathbf{h}$ is unknown and is being estimated.

In all compared algorithms, the parameters were optimized up to a feasible extent, based on the suggestions of the authors as well as testing a large set of different values and the best obtained results are presented here. For the non-blind experiments we, of course, skip step *1.d*) of the proposed algorithm, since the blur is assumed to be known. The algorithm is terminated when the convergence criterion $\|\mathbf{x}^k - \mathbf{x}^{k-1}\|/\|\mathbf{x}^{k-1}\| < 10^{-3}$ is satisfied. Additionally, the initial blur estimate $\mathbf{h}^{1,1}$ is set to a Gaussian of variance 2.

We conduct a series of experiments for compressive ratios 0.1 to 1 with step 0.1 for all algorithms and the average performance is reported. We test each algorithm under different blur degradations and noise realizations. Specifically, we degrade each of the tested images with Gaussian PSFs of variance 5 and 9 and add Gaussian noise to the CS measurements so that the signal to noise ratio (SNR) becomes 40 dB (additional cases were tested but are not presented here due to space limitations). A performance comparison is presented in Figure 1 which documents that our approach outperforms the non-blind results of state-of-the art methods, as well as the blind results of our previous framework, for the majority of the tested scenarios. For the blind case, our algorithm performs reasonably well, achieving PSNRs close to the non-blind case, especially for the "Shepp-Logan" image.

Figure 2 depicts the tested images, their blurry degraded versions with a Gaussian PSF of variance 5, and examples of blind and non-blind reconstructions/restorations for different CS ratios when SNR = 40 dB. The presented images correspond to the highest achieved PSNR out of the tested cases for each CS ratio. It is apparent that images of high visual quality are obtained following the proposed framework.

5. CONCLUSIONS

In this paper, we proposed a novel variational Bayesian BID framework for imaging systems based on the principles of CS and provided simultaneous estimation rules for all modeling parameters, the image, and the blur. Moreover, we demonstrated how to adopt and generalize the alternating direction method of multipliers, often utilized to solve constrained optimization problems, within the variational Bayesian approach. Finally, the proposed framework is general and can easily be adapted to different BID approaches that utilize alternative regularization terms for the image and blur. In the experimental section we showed that the proposed approach has advantages over existing CS algorithms.

6. REFERENCES

[1] C. Hegde and R. G. Baraniuk, "Compressive sensing of streams of pulses," in *Proc. 47th annual Allerton Conf. Communication, control, and computing*. 2009, Allerton'09, pp. 44–51, IEEE Press.

[2] C. Hegde and R. G. Baraniuk, "Sampling and recovery of pulse streams," *IEEE Trans. Signal Process.*, vol. 59, no. 4, pp. 1505–1517, Apr. 2011.

[3] B. Amizic, L. Spinoulas, R. Molina, and A. K. Katsaggelos, "Compressive sampling with unknown blurring function: Application to passive millimeter-wave imaging," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Oct. 2012, pp. 925–928.

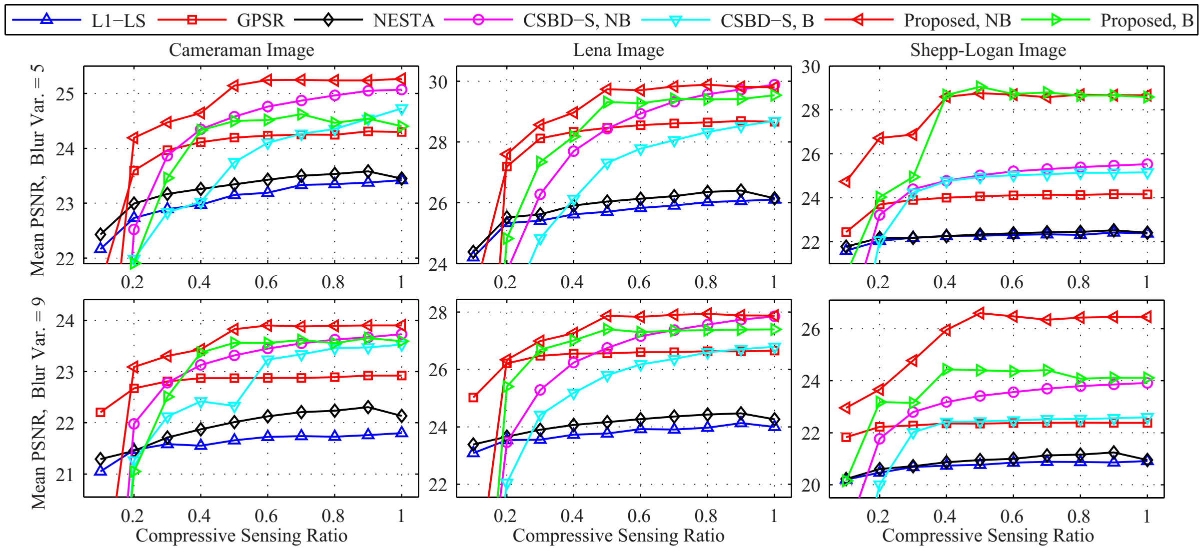

Fig. 1: Average performance comparison of the non-blind (NB) and blind (B) versions of the proposed approach with a series of state-of-the-art NB algorithms, namely, *l1-ls* [12], *GPSR* [13], *NESTA* [14] and our previous framework in [3] (denoted as *CSBD-S*).

Fig. 2: Example reconstructions: For each image, 1^{st} row represents the original; the degraded image with a Gaussian PSF of variance 5 and the non-blind restoration result for SNR = 40 dB and 70% of CS measurements, from left to right, respectively. 2^{nd} row represents the blind deconvolution results, using the proposed framework, for 70%, 50% and 30% of CS measurements, from left to right, respectively.

[4] E. J. Candès and J. Romberg, "Practical Signal Recovery from Random Projections.," *Wavelet Applications in Signal and Image Processing XI, Proc. SPIE Conf.*, vol. 5914, 2005.

[5] V. Cevher, "Learning with compressible priors.," in *NIPS*. 2009, pp. 261–269, Curran Associates, Inc.

[6] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[7] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," in *SIGGRAPH '07*. 2007, ACM.

[8] B. Amizic, R. Molina, and A. K. Katsaggelos, "Sparse bayesian blind image deconvolution with parameter estimation," *EURASIP J. Image and Video Processing*, vol. 20, 2012.

[9] C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

[10] S. Babacan, R. Molina, and A. Katsaggelos, "Parameter estimation in TV image restoration using variational distribution approximation," *IEEE Trans. Image Process.*, vol. 17, no. 3, pp. 326–339, 2008.

[11] J. Bioucas-Dias, M. Figueiredo, and J. Oliveira, "Adaptive total-variation image deconvolution: A majorization-minimization approach," in *Proc. EUSIPCO'2006*, Florence, Italy, Sept. 2006.

[12] S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An interior-point method for large-scale l1-regularized least squares," *IEEE J. Sel. Topics in Signal Process.*, vol. 1, no. 4, pp. 606–617, 2007.

[13] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE J. of Sel. Topics Signal Process.*, vol. 1, no. 4, pp. 586 –597, 2007.

[14] S. Becker, J. Bobin, and E. J. Candés, "Nesta: A fast and accurate first-order method for sparse recovery," *SIAM J. Imaging Sciences*, vol. 4, no. 1, pp. 1–39, 2011.