DETAILS ON *CSA-SBL*:
AN ALGORITHM FOR SPARSE BAYESIAN LEARNING BOOSTED BY PARTIAL ERRONEOUS SUPPORT KNOWLEDGE

Mohammad Shekaramiz, Todd K. Moon, and Jacob H. Gunther

Electrical and Computer Engineering Department and Information Dynamics Laboratory
Utah State University
{mohammad.shekaramiz@aggiemail.usu.edu, todd.moon@usu.edu, jake.gunther@usu.edu}

ABSTRACT

This report provides details on CSA-SBL(VB) algorithm for the recovery of sparse signals with unknown clustering pattern. More specifically, we deal with the recovery of sparse signals with unknown clustering pattern in the case of having partial erroneous prior knowledge on the supports of the signal. In [1], we provided a modified sparse Bayesian learning model to incorporate prior knowledge and simultaneously learn the unknown clustering pattern. For this purpose, we added one more layer to support-aided sparse Bayesian learning algorithm (SA-SBL) that was proposed in [2]. This layer adds a prior on the shape parameters of Gamma distributions, those modeled to account for the precision of the solution elements. We made the shape parameters depend on the total variations on the estimated supports of the solution. The inference technique for implementing this algorithm is variational Bayes (VB). Part of this work has been published in [1, 3].

Keywords— Sparse Bayesian learning (SBL), single measurement vector (SMV), compressive sensing, clustered pattern, erroneous support aided.

1. INTRODUCTION

Compressive sensing (CS) provides tools to represent a sparse or compressible signal from a small set of non-adaptive linear measurements [4]. In linear CS, the high dimensional signal $\mathbf{x} \in \mathbb{R}^N$ is modeled by the linear equation $\mathbf{y} = A\mathbf{x} + \mathbf{e}$, where $A \in \mathbb{R}^{M \times N}$ is a wide sensing matrix with $M \ll N$. The case where the sensing matrix is known has been referred to as single measurement vector (SMV) problem [5]. In the CS context, it is assumed that $\mathbf{x}$ is sparse (has few non-zero elements) under some proper basis. Besides the sparsity, in some practical applications the nonzero entries of the sparse signal $\mathbf{x}$ may appear in clusters. This feature has been referred to as clustered-pattern or block-sparsity in the literature [6, 7]. Moreover, there exist cases where a partial erroneous support set of the solution is available as a prior knowledge. This type of information may be obtained from either of the two cases below. It can be the estimate of the supports inferred from the set of measurements taken from a phenomenon of interest at the last time instant. Here, we investigate the sparse recovery problem of signals with unknown clustering pattern for the case where some prior information on the supports of the solution is available. More specifically, we assume that we are provided with a partial erroneous support set of the solution [2].

In case of having prior knowledge on the supports of the solution, some algorithms such as MBPDN and SA-SBL have been recently proposed [2, 8]. MBPDN algorithm is a modified version of basis pursuit de-noising algorithm for the case where a subset of true support set is available [8]. Recently, it has been shown [2] that MBPDN is sensitive to the accuracy of the prior knowledge on the available support set. In [2], Fang et al. proposed a modified version of the conventional sparse Bayesian learning model for the purpose of using prior information on the support set in order to obtain better estimate of the true underlying sparse signal. The conventional SBL algorithm considers a Gaussian-inverse-Gamma distribution on the elements of the solution vector $\mathbf{x}$. In [2], one more layer was added to the conventional SBL. This layer incorporates a prior on the rate parameter of the Gamma distribution to take advantage of the available support knowledge of the solution. Following the same notations as was used in [2], suppose that $\mathcal{T}$ is the set of all the true supports in the solution and $\mathcal{S} \subset \mathcal{T}$ is a subset of true supports that is available. Furthermore, assume that $\mathcal{E} \subset \mathcal{T}^c$ contains the error subset that is incorrectly considered as a part of available true supports. Notice that $\mathcal{T} \cup \mathcal{T}^c = \{1, 2, \ldots, N\}$ and $\mathcal{P} = \mathcal{S} \cup \mathcal{E}$, where $\mathcal{P}$ is the erroneous support set that is available to us. As discussed earlier, one can think of $\mathcal{P}$ as the support set of the previous column of the solution matrix in the MMV problem.

Below, we briefly describe the priors that were considered in [2]. Each element x_n of the solution was assumed to be drawn *i.i.d.* from zero mean Gaussian distribution with corresponding precision α_n, i.e.,

$$x_n \sim \mathcal{N}(x_n; 0, \alpha_n^{-1}), \ \forall n = 1, \ldots, N, \tag{1}$$

where the precisions are random variables defined as

$$\alpha_n \sim \text{Gamma}(\alpha_n; a, b_n), \ \forall n = 1, \ldots, N, \tag{2}$$

where the shape parameter is fixed to $a = 10^{-10}$. In order to incorporate the available and probably erroneous support knowledge, Fang et al. [2] defined the two following cases on the rate parameters of the Gamma distributions

$$\begin{cases} b_n \sim \text{Gamma}(p, q), & \text{if } n \in \mathcal{P} \\ b_n = 10^{-15}, & \text{otherwise} \end{cases} \tag{3}$$

This means that only when the index n belongs to the set $\mathcal{P}$ the corresponding precision α_n will be governed by another Gamma distribution with hyper-parameters p and q. In this case, SA-SBL algorithm was proposed in [2].

The CSA-SBL algorithm is essentially a modified version of SA-SBL, in which we also account for the unknown clustering pattern that may exist in the original signal. For this purpose, we incorporate the measure of clumpiness over the supports of the solution $(\Sigma\Delta)$ proposed in [9] into SA-SBL algorithm. The main difference between CSA-SBL and SA-SBL is that we further put a prior on the shape parameter of the Gamma distribution defined in (2) while it was set to a constant in SA-SBL [2]. Specifically, we impose that the shape parameter is to be controlled by the estimated measure of contiguity in the supports of the solution i.e., total variation on the supports of the solution. Based on the simulations, we show that this modification improves the overall performance in estimating the supports of the solution.

2. A BRIEF REVIEW ON VARIATIONAL BAYES INFERENCE

Variational Bayes is an effective approach to approximate intractable integrals that occur in Bayesian inference. VB provides analytical approximation to the posterior distributions of the parameters and hidden variables of statistical models using a lower bound on the marginal likelihood of the observations. VB is essentially an extension of expectation-maximization (EM) algorithm [10, 11]. Suppose there is a probabilistic model with parameters Θ, hidden variables $\mathbf{x}$, and a set of observations denoted by $\mathbf{y}$. Then, the approximation to the joint density $p(\mathbf{x}, \Theta|\mathbf{y})$ can be represented by $p(\mathbf{x}, \Theta|\mathbf{y}) \approx q_x(\mathbf{x})q_\theta(\Theta)$. In the ideal case, we desire to select $Q_{x,\theta}(\mathbf{x}, \Theta) = q_x(\mathbf{x})q_\theta(\Theta)$ to be as close as possible to $p(\mathbf{x}, \Theta|\mathbf{y})$. Since computing the normalization factor (probability of the observation) is intractable, we write the logarithm of the evidence in terms of the integral of the joint probability $p(\mathbf{x}, \Theta, \mathbf{y})$ and then incorporate $Q_{x,\theta}(\mathbf{x}, \Theta)$ into the integrand. Using Jensen's inequality, we obtain a lower bound on the logarithm of evidence. The problem turns into maximizing this lower bound to make $Q_{x,\theta}(\mathbf{x}, \Theta)$ close to $p(\mathbf{x}, \Theta|\mathbf{y})$. The lower bound on the model log-marginal likelihood can be iteratively optimized by the following updates [10, 12]

$$q_x^{[t+1]}(\mathbf{x}) \propto \exp\left\{ <\log p(\mathbf{x}, \mathbf{y}|\Theta)) >_{q_\theta^{[t]}} \right\} \tag{4}$$

$$q_\Theta^{[t+1]}(\Theta) \propto p(\Theta) \exp\left\{ <\log p(\mathbf{x}, \mathbf{y}|\Theta) >_{q_x^{[t]}} \right\}, \tag{5}$$

where the term $<p>_q$ denotes the expectation of p under the density q.

3. CSA-SBL ALGORITHM

In this section we describe our modified version of the conventional SBL algorithm to solve for $\mathbf{x}$ in the SMV problem defined by $\mathbf{y} = A\mathbf{x} + \mathbf{e}$. It is assumed that an erroneous support set $\mathcal{P} = \mathcal{S} \cup \mathcal{E}$ is available, where $\mathcal{S}$ is a subset of the true supports of the solution and $\mathcal{E}$ is a set of incorrectly considered supports. The partition $\mathcal{S}$ and $\mathcal{E}$ in $\mathcal{P}$ is assumed unknown [2]. In order to account for the clustering pattern that may exist in the solution, we borrow the measure of clumpiness from [9, 13], which is defined as

$$(\Sigma\Delta)_{(\text{support of } \mathbf{x})} = \sum_{n=2}^{N} |b(x_n, T) - b(x_{n-1}, T)|, \tag{6}$$

where T is a predetermined threshold. The function $b(\cdot, \cdot)$ in (6) returns a binary value and is defined as follows

$$b(x_n, T) = \begin{cases} 1 & \text{if } |x_n| > T \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

The more clustered the solution becomes, the lower value $(\Sigma\Delta)$ in (6) will possess. Based on (7), the entries of $\mathbf{x}$ with the amplitude less than the threshold T are deemed to zero and their corresponding index will not be considered as supports of the solution. This is due to the fact that the elements do not have significant contribution in our measurements. We experimentally set $T = 10^{-6}$.

Below we describe the prior distributions that we consider in our hierarchical Bayesian model. Similar to SBL and for the purpose of promoting sparsity in the solution, we assume that the elements of the solution are drawn *i.i.d.* from a zero-mean Gaussian distribution as follows

$$x_n \sim \mathcal{N}(x_n; 0, \alpha_n^{-1}), \ \forall n = 1, \ldots, N, \tag{Revisiting (1)}$$

where the precision α_n is modeled as

$$\alpha_n \sim \text{Gamma}(\alpha_n; a_n, b_n), \ \forall n = 1, \ldots, N. \tag{8}$$

Unlike (2), we do not assign the same shape parameter to the precisions $\alpha_n, \ \forall n = 1, \ldots, N$ in our model.

In order to incorporate the available erroneous support knowledge, we use the same model as defined [2] for the rate parameters defined in (8).

$$\begin{cases} b_n \sim \text{Gamma}(b_n; p, q), & \text{if } n \in \mathcal{P} \\ b_n = 10^{-15}, & \text{otherwise} \end{cases}, \tag{Revisiting (3)}$$

where $p = q = 0.1$ as suggested in [2].

The reason for having different shape parameters for the precisions α_n in (8) is to promote the clustered pattern in the support set of the solution, where the pattern is learned the via measure of total variation on the supports defined in (6). In other words, we let each shape parameter a_n be controlled via the estimated $(\Sigma\Delta)$. For this purpose, we add another hyper-prior to our model as follows

$$a_n \sim \text{Gamma}(a_n; g_n, h), \ \forall n = 1, \ldots, N, \tag{9}$$

where

$$g_n := \theta \exp\left\{\left((\Sigma\Delta) - (\Sigma\Delta)_{n,0}\right)\right\}, \tag{10}$$

where $(\Sigma\Delta)$ is the initial measure of the clumpiness based on the available erroneous support set $\mathcal{P}$ and is computed from (6), and $(\Sigma\Delta)_{n,0}$ is the measure of clumpiness when forcing $x_n = 0$. In (10), θ is an emphasizing parameter on the amount of clumpiness over the supports of the solution and one can make it depend on the ratio M/N. This means that when the number of measurements is very low, we may not wish to emphasize on the clustered solutions. Otherwise, the algorithm may also remove some of the true supports in the set $\mathcal{P}$ due to the small number of measurements and the lack of information on the full true support set.

Remark: As a default setting, we suggest $h = 1$ but in general, 'h' is selected based on the belief on the maximum permissible amplitude of the elements of $\mathbf{x}$.

Finally, the prior on the noise is defined as

$$\sigma^2 \sim \mathcal{N}(0, \gamma^{-1}), \gamma \sim \text{Gamma}(c, d), \tag{11}$$

where we set $c = d = 10^{-4}$ as suggested in [2].

According to the above prior distributions, the joint probability distribution of our model becomes

$$P(\mathbf{y}, \mathbf{x}, \boldsymbol{\alpha}, \mathbf{a}, \mathbf{b}, \sigma^2, \gamma) \propto P(\mathbf{y}|\mathbf{x}, \sigma^2 I_M) P(\mathbf{x}; 0, \operatorname{diag}\{\boldsymbol{\alpha}\}^{-1}) \times$$
$$P(\boldsymbol{\alpha}; \mathbf{a}, \mathbf{b}) P(\mathbf{a}; \mathbf{g}, h) p(\sigma^2; 0, \gamma^{-1}) p(\gamma; c, d) \prod_{n \in \mathcal{P}} P(b_n; p, q), \tag{12}$$

where the measurement noise is assumed to be $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_M)$.

According to (12), the marginalized posterior distributions for the variables of interest can be represented as follows. In these descriptions, conditioning on $-$, as in $(x_n|-)$, denotes the inference on x_n conditioning upon all relevant variables and the observations.

- $P(\mathbf{x}|-) \propto P(\mathbf{y}|\mathbf{x}, \sigma^2 I_M) P(\mathbf{x}|D^{-1})$. Therefore,

$$(\mathbf{x}|-) \sim \mathcal{N}(\tilde{\mathbf{x}}, \Sigma_{\tilde{x}}), \tag{13}$$

  where

$$\tilde{\mathbf{x}} = \gamma \Sigma_{\tilde{x}} A^T \mathbf{y}$$
$$\Sigma_{\tilde{x}} = (\gamma A^T A + D)^{-1},$$

  where D is a diagonal matrix with $\boldsymbol{\alpha}$ as its main diagonal i.e., $[D]_{n,n} = \alpha_n$.

- $P(\boldsymbol{\alpha}|-) \propto P(\mathbf{x}|D^{-1}) P(\boldsymbol{\alpha}; \mathbf{a}, \mathbf{b})$. Therefore,

$$(\alpha_n|-) \sim \begin{cases} \operatorname{Gamma}(\bar{a}_n, b_n + \frac{\tilde{x}_n^2 + \sigma_{\tilde{x}_n}^2}{2}), & \text{if } n \in \mathcal{P} \\ \operatorname{Gamma}(\bar{a}_n, 10^{-15} + \frac{\tilde{x}_n^2 + \sigma_{\tilde{x}_n}^2}{2}), & \text{otherwise} \end{cases}, \tag{14}$$

  In the above equation, $\bar{a}_n := a_n + \frac{1}{2}$ and $\sigma_{\tilde{x}_n}^2 := \alpha_n^{-1}$.

- $P(a_n|-) \propto P(\alpha_n; a_n, b_n) P(a_n|g_n, h), \forall n = 1, \ldots, N$. Hence

$$(a_n|-) \sim \operatorname{Gamma}(g_n, h + \log \alpha_n), \ \forall n = 1, \ldots, N, \tag{15}$$

  where $g_n = \theta \exp\{(\hat{\Sigma \Delta})^{[k]} - (\hat{\Sigma \Delta})_{n,0}^{[k]}\}$ and k denotes the kth iteration and $(\hat{\Sigma \Delta})$ is the estimated measure of clumpiness.

- $P(b_n|-) \propto P(\alpha_n; a_n, b_n) P(b_n; p, q), \forall n \in \mathcal{P}$. As a result,

$$(b_n|-) \sim \operatorname{Gamma}(p, q + \alpha_n), \ \forall n \in \mathcal{P}. \tag{16}$$

- Finally, $P(\gamma|-) \propto P(\mathbf{y}|\mathbf{x}, \gamma) P(\gamma; c, d)$ and therefore,

$$(\gamma|-) \sim \operatorname{Gamma}\left(c + \frac{M}{2}, d + \frac{1}{2}\left(\|\mathbf{y} - A\tilde{\mathbf{x}}\|_2^2 + \operatorname{tr}(A^T A \Sigma_{\tilde{x}})\right)\right). \tag{17}$$

4. APPENDIX

In this section, we provide details on the update rules of the variables in the CSA-SBL modeling and algorithm. Using the VB technique, the update rule of the variables and parameters of the model can be simplified as follows.

- Update rule for the solution vector '$\mathbf{x}$'

$$q_x(x) \sim \exp\left\{ < \log\{p(\mathbf{x}, \mathbf{y}|\theta)\} >_{q_\theta} \right\},$$

which yields to

$$q_x(x) \propto e^{<\log\{p(\mathbf{x}|\theta)\}>_{q_\theta}} e^{<\log\{p(\mathbf{y}|\mathbf{x},\theta)\}>_{q_\theta}}$$

$$\propto e^{<\log\{p(\mathbf{x}|D^{-1})\}>_{q_\alpha}} e^{<\log\{p(\mathbf{y}|\mathbf{x},\gamma^{-1}I_M)\}>_{q_\theta}}$$

$$\propto e^{<\log\{|D|^{\frac{1}{2}} e^{-\frac{1}{2}\mathbf{x}^T D\mathbf{x}}\}>_{q_\alpha}} e^{<\log\{\gamma^{\frac{M}{2}} e^{-\frac{1}{2}\gamma\|\mathbf{y}-A\mathbf{x}\|_2^2}\}>_{q_\gamma}}$$

$$\propto e^{-\frac{1}{2}<\mathbf{x}^T D\mathbf{x}>_{q_\alpha}} e^{-\frac{1}{2}\tilde{\gamma}\|\mathbf{y}-A\mathbf{x}\|_2^2}$$

$$\propto e^{-\frac{1}{2}<\mathbf{x}^T D\mathbf{x}>_{q_\alpha}} e^{-\frac{1}{2}\tilde{\gamma}(\mathbf{x}^T A^T A\mathbf{x} - 2\mathbf{x}^T A^T \mathbf{y})}$$

$$\propto e^{\left\{-\frac{1}{2}\left(\mathbf{x}^T \tilde{D}\mathbf{x} + \tilde{\gamma}(\mathbf{x}^T A^T A\mathbf{x} - 2\mathbf{x}^T A^T \mathbf{y})\right)\right\}}$$

$$\propto e^{\left\{-\frac{1}{2}\left(\mathbf{x}^T(\tilde{D}+\tilde{\gamma}A^T A)\mathbf{x} - 2\tilde{\gamma}\mathbf{x}^T A^T \mathbf{y}\right)\right\}},$$

where D is a diagonal matrix and is defined as $D := \text{diag}\{\alpha_1, \dots, \alpha_N\}$. Therefore,

$$q_x(x) \propto e^{\left\{-\frac{1}{2}\left(\left(\mathbf{x}-(\tilde{D}+\tilde{\gamma}A^T A)^{-1}\tilde{\gamma}A^T \mathbf{y}\right)(\tilde{D}+\tilde{\gamma}A^T A)(\star)\right)\right\}}$$

Therefore, $q_x(x) \sim \mathcal{N}(\tilde{\mathbf{x}}, \Sigma_{\tilde{x}})$, where

$$\Sigma_{\tilde{x}} = (\tilde{D} + \tilde{\gamma}A^T A)^{-1}$$

$$\tilde{\mathbf{x}} = \tilde{\gamma}\Sigma_{\tilde{x}}A^T \mathbf{y}$$

- Update rule for 'γ' (noise precision)

$$q(\gamma) \sim p(\gamma; c, d) \exp < \log p(\mathbf{y}|\mathbf{x}, \gamma^{-1}I_M) >_{q_x},$$

which yields to

$$q(\gamma) \propto \gamma^{c-1} e^{-d\gamma} e^{\left\{<\log\{\gamma^{\frac{M}{2}} e^{-\frac{1}{2}\gamma\|\mathbf{y}-A\mathbf{x}\|_2^2}\}>_{q_x}\right\}}$$

$$\propto \gamma^{c-1} e^{-d\gamma} \gamma^{\frac{M}{2}} e^{\left\{-\frac{1}{2}\gamma<\|\mathbf{y}-A\mathbf{x}\|_2^2>_{q_x}\right\}}$$

$$\propto \gamma^{(c+\frac{M}{2})-1} e^{\left\{-(d+\frac{1}{2}<\|\mathbf{y}-A\mathbf{x}\|_2^2>_{q_x})\right\}}$$

Remark:

$$< \|\mathbf{y} - A\mathbf{x}\|_2^2 >_{q_x} = < \mathbf{y}^T\mathbf{y} - 2\mathbf{x}^T A^T \mathbf{y} + \mathbf{x}^T A^T A\mathbf{x} >_{q_x}$$

$$= \mathbf{y}^T\mathbf{y} - 2 < \mathbf{x} >_{q_x}^T A^T \mathbf{y} + < \mathbf{x}^T A^T A\mathbf{x} >_{q_x}$$

Define $\tilde{\mathbf{x}} := < x >_{q_x}$.

Notice that

$$< \mathbf{x}^T A^T A \mathbf{x} >_{q_x} = \mathrm{tr}\left(< \mathbf{x}^T A^T A \mathbf{x} >_{q_x}\right)$$
$$= \mathrm{tr}\left(< \mathbf{x}\mathbf{x}^T >_{q_x} A^T A\right)$$
$$= \mathrm{tr}\left((\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T + \Sigma_{\tilde{x}})A^T A\right).$$

Therefore,

$$< \|\mathbf{y} - A\mathbf{x}\|_2^2 >_{q_x} = \mathbf{y}^T \mathbf{y} - 2\tilde{\mathbf{x}}^T A^T \mathbf{y} + \mathrm{tr}\left((\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T + \Sigma_{\tilde{x}})A^T A\right)$$

Therefore,

$$q(\gamma) \propto \gamma^{(c+\frac{M}{2})-1} e^{-(d+\frac{1}{2}\tilde{\Phi})},$$

where

$$\tilde{\Phi} = \mathbf{y}^T \mathbf{y} - 2\tilde{\mathbf{x}}^T A^T \mathbf{y} + \mathrm{tr}\left((\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T + \Sigma_{\tilde{x}})A^T A\right)$$
$$= \|\mathbf{y} - A\tilde{\mathbf{x}}\|_2^2 + \mathrm{tr}\left(A^T A \Sigma_{\tilde{x}}\right),$$

which yields to

$$q(\gamma) \sim \Gamma(c + \frac{M}{2}, d + \frac{1}{2}\tilde{\Phi}).$$

Finally, the update rule for the noise precision becomes

$$\tilde{\gamma} = \frac{c + \frac{M}{2}}{d + \frac{1}{2}\tilde{\Phi}}.$$

- Update rule for the parameter 'a_n', $\forall n \in \mathcal{P}$

$$q(a_n) \sim p(a_n; g_n, h) \exp\left\{< \log p(\alpha_n; a_n, b_n) >_{q_{\alpha_n} q_{b_n}}\right\},$$

which yields to

$$q(a_n) \propto a_n^{g_n - 1} e^{-h a_n} e^{\left\{< \log\left\{\alpha_n^{a_n - 1} e^{-b_n \alpha_n}\right\}>_{q_{\alpha_n} q_{b_n}}\right\}}$$
$$\propto a_n^{g_n - 1} e^{-h a_n} e^{\left\{(a_n - 1)< \log \alpha_n >_{q_{\alpha_n}} -<b_n>_{q_{b_n}} \alpha_n\right\}}$$
$$\propto a_n^{g_n - 1} e^{-h a_n} e^{< \log \alpha_n >_{q_{\alpha_n}} a_n}$$
$$\propto a_n^{g_n - 1} e^{-(h - < \log \alpha_n >_{q_{\alpha_n}})a_n}$$

Remark: $< \log \alpha_n >_{q_{\alpha_n}} = \Psi(\bar{a}_n) - \log \bar{b}_n$, where

$$\begin{cases} \bar{a}_n = a_n + \frac{1}{2} \\ \bar{b}_n = b_n + \frac{1}{2}(\tilde{x}_n^2 + \sigma_{\tilde{x}_n}^2), \forall n \in \mathcal{P} \\ \bar{b}_n = 10^{-15}, \forall n \notin \mathcal{P}. \end{cases}$$

Therefore,

$$q(a_n) \propto a_n^{g_n - 1} e^{-\left(h + \log \bar{b}_n - \Psi(\bar{a}_n)\right)a_n},$$

or equivalently

$$q(a_n) \sim \Gamma\left(g_n, h + \log \bar{b}_n - \Psi(\bar{a}_n)\right),$$

which yields to the below update rule

$$\tilde{a}_n = \frac{g_n}{h + \log \bar{b}_n - \Psi(\bar{a}_n)}.$$

- Update rule for the parameter set '$\boldsymbol{\alpha}$' (precision on the solution vector $\mathbf{x}$)

$$q_{\alpha_n}(\alpha_n) \sim p(\alpha_n; \tilde{a}_n, \tilde{b}_n) \exp\left\{< \log\{p(x_n|0, \alpha_n^{-1})\} >_{q_{x_n}}\right\},$$

which yields to

$$q_{\alpha_n}(\alpha_n) \propto \alpha_n^{\tilde{a}_n - 1} e^{-\tilde{b}_n \alpha_n} e^{\left\{<\log\{\alpha_n^{\frac{1}{2}} e^{-\frac{1}{2}\alpha_n x_n^2}\} >_{q_{x_n}}\right\}}$$

$$\propto \alpha_n^{\tilde{a}_n - 1} e^{-\tilde{b}_n \alpha_n} \alpha_n^{\frac{1}{2}} e^{-\frac{1}{2}\alpha_n <x_n^2>_{q_{x_n}}}$$

$$\propto \alpha_n^{(\tilde{a}_n + \frac{1}{2}) - 1} e^{-\tilde{b}_n \alpha_n} e^{-\frac{1}{2}\alpha_n(\tilde{x}_n^2 + \sigma_{\tilde{x}_n}^2)}$$

$$\propto \alpha_n^{(\tilde{a}_n + \frac{1}{2}) - 1} e^{-\left(\tilde{b}_n + \frac{1}{2}(\tilde{x}_n^2 + \sigma_{\tilde{x}_n}^2)\right)\alpha_n}$$

Therefore,

$$q_{\alpha_n}(\alpha_n) \sim \Gamma\left(\tilde{a}_n + \frac{1}{2}, \tilde{b}_n + \frac{1}{2}(\tilde{x}_n^2 + \sigma_{\tilde{x}_n}^2)\right), \ \forall n$$

and the update rule for the precision α_n becomes

$$\tilde{\alpha}_n = \frac{\tilde{a}_n + \frac{1}{2}}{\tilde{b}_n + \frac{1}{2}(\tilde{x}_n^2 + \sigma_{\tilde{x}_n}^2)}, \forall n.$$

- Update rule for the parameter set '$\mathbf{b}$', $\forall n \in \mathcal{P}$

$$q_{b_n}(b_n) \sim p(b_n; p, q) \exp\left\{< \log\{p(\alpha_n; a_n, b_n)\} >_{q_{\alpha_n} q_{a_n}}\right\},$$

which yields to

$$q_{b_n}(b_n) \propto b_n^{p-1} e^{-q b_n} e^{\left\{<\log\{\alpha_n^{a_n - 1} e^{-b_n \alpha_n}\} >_{q_{\alpha_n} q_{a_n}}\right\}}$$

$$\propto b_n^{p-1} e^{-q b_n} e^{\left\{(\tilde{a}_n - 1)<\log \alpha_n>_{q_{\alpha_n}} - \tilde{\alpha}_n b_n\right\}}$$

$$\propto b_n^{p-1} e^{-q b_n} e^{-\tilde{\alpha}_n b_n}$$

$$\propto b_n^{p-1} e^{-(q + \tilde{\alpha}_n) b_n}$$

Therefore,

$$q_{b_n}(b_n) \sim \Gamma(p, q + \tilde{\alpha}_n), \forall n \in \mathcal{P},$$

and the update rule for 'b_n' becomes

$$\tilde{b}_n = \frac{p}{q + \tilde{\alpha}_n}, \forall n \in \mathcal{P}.$$

5. REFERENCES

[1] M. Shekaramiz, T. Moon, and J. Gunther, "Sparse Bayesian learning boosted by partial erroneous support knowledge," in *50th Asilomar Conf. on Sig., Syst. and Compt.*, pp. 389–393, Nov. 2016.

[2] J. Fang, Y. Shen, F. Li, H. Li, and Z. Chen, "Support knowledge-aided sparse Bayesian learning for compressed sensing," in *IEEE Int. Conf. on Acoust., Speech and Sig. Proc. (ICASSP)*, pp. 3786–3790, 2015.

[3] M. Shekaramiz, *Sparse Signal Recovery Based on Compressive Sensing and Exploration Using Multiple Mobile Sensors*. PhD Dissertation, Utah State University, Digitalcommons, 2018.

[4] D. Donoho, "Compressed sensing," *IEEE Trans. Info. Th.*, vol. 52, pp. 1289–1306, 2006.

[5] M. Mishali and Y. C. Eldar, "Reduce and boost: Recoverying arbitrary sets of jointly sparse vectors," *IEEE Trans. Sig. Proc.*, vol. 56, no. 10, pp. 4692–4702, 2008.

[6] Z. Zhang and B. Rao, "Recovery of block sparse signals using the framework of block sparse Bayesian learning," in *IEEE Int. Conf. on Acoust., Speech and Sig. Proc. (ICASSP)*, pp. 3345–3348, 2012.

[7] J. Fang, Y. Shen, H. Li, and P. Wang, "Pattern-coupled sparse Bayesian learning for recovery of block-sparse signals," *IEEE Trans. Sig. Proc.*, vol. 63, no. 2, pp. 360–372, 2015.

[8] N. Vaswani and W. Lu, "Modified-CS: Modifying compressive sensing for problems with partially known support," *IEEE Trans. Sig. Proc.*, vol. 58, no. 9, pp. 4595–4607, 2010.

[9] M. Shekaramiz, T. Moon, and J. Gunther, "On the block-sparse solution of single measurement vectors," in *49th Asilomar Conf. on Sig., Syst. and Compt.*, pp. 508–512, Nov. 2015.

[10] M. Beal, *Variational Algorithms for Approximate Bayesian Inference*. PhD dissertation, University College London, 2003.

[11] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Sig. Proc. Mag.*, vol. 25, no. 6, pp. 131–142, 2008.

[12] C. Fox and S. Roberts, "A tutorial on variational Bayesian inference," *Artificial Intelligence Review*, vol. 38, no. 2, pp. 85–95, 2011.

[13] M. Shekaramiz, T. Moon, and J. Gunther, "Hierarchical Bayesian approach for jointly-sparse solution of multiple-measurement vectors," in *48th Asilomar Conf. on Sig., Syst. and Compt.*, pp. 1962–1966, Nov. 2014.