Low Complexity Sparse Bayesian Learning for Channel Estimation Using Generalized Mean Field

Pedersen, Niels Lovmand; Manchón, Carles Navarro; Fleury, Bernard Henri

Published in:
European Wireless 2014; 20th European Wireless Conference; Proceedings of

Publication date:
2014

Document Version
Accepted author manuscript, peer reviewed version

Low Complexity Sparse Bayesian Learning for Channel Estimation Using Generalized Mean Field

Niels Lovmand Pedersen, Carles Navarro Manchón and Bernard Henri Fleury
Department of Electronic Systems, Aalborg University
Niels Jernes Vej 12, DK-9220 Aalborg, Denmark, Email: {nlp,cnm,bfl}@es.aau.dk

***Abstract*—We derive low complexity versions of a wide range of algorithms for sparse Bayesian learning (SBL) in underdetermined linear systems. The proposed algorithms are obtained by applying the generalized mean field (GMF) inference framework to a generic SBL probabilistic model. In the GMF framework, we constrain the auxiliary function approximating the posterior probability density function of the unknown variables to factorize over disjoint groups of contiguous entries in the sparse vector - the size of these groups dictates the degree of complexity reduction. The original high-complexity algorithms correspond to the particular case when all the entries of the sparse vector are assigned to one single group. Numerical investigations are conducted for both a generic compressive sensing application and for channel estimation in an orthogonal frequency-division multiplexing receiver. They show that, by choosing small group sizes, the resulting algorithms perform nearly as well as their original counterparts but with much less computational complexity.**

I. Introduction

Compressive sensing and sparse signal representation have proven to be very useful tools in a large variety of engineering areas. One application in wireless communications, which we address in this paper, is the estimation of the radio channel by exploiting its inherent sparse nature. The high practicability of compressive sensing has sparked the development of a growing number of techniques for recovering sparse signals in underdetermined linear systems. The classical signal model assumes that a vector $\boldsymbol{y}$ consisting of M observations is obtained from the $N > M$ dimensional sparse weight vector $\boldsymbol{w}$ according to

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{w} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{\Phi} = [\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_N]$ is referred to as the $M \times N$ dictionary matrix and $\boldsymbol{n}$ is additive white Gaussian noise with covariance matrix $\lambda^{-1}\boldsymbol{I}$. The vector $\boldsymbol{w}$ is K-sparse in the canonical basis and is assumed to have statistically independent nonzero entries. Due to $N > M$, classical (penalized) least-squares estimates will produce non-sparse solutions for $\boldsymbol{w}$. As a result, many convex [1], [2], greedy [3], and Bayesian methods aiming at finding sparse estimates of the weight vector have been proposed in the literature in recent years. In this paper, we focus on methods based on sparse Bayesian learning (SBL).

One popular SBL algorithm is the relevance vector machine (RVM) [4]. Recovering $\boldsymbol{w}$ using RVM is, nevertheless, of substantial computational complexity and is often disregarded even though the performance is on par with many state-of-the-art algorithms. In order to lower the computational requirements of RVM, a greedy-based inference scheme is proposed in [5] and later applied in [6], [7].

In this paper, we develop iterative, low complexity SBL algorithms, which have a computational complexity per algorithmic iteration that is lower than that of the methods in [5]–[7] while being non-greedy. The inference framework is valid for the estimation of real- and complex-valued signals.

Our approach is based on generalized mean field (GMF) inference [8]–[10]. Roughly speaking, GMF approximates the posterior probability density function (pdf) of a set of unknown variables with an auxiliary function, which is constrained to factorize over groups of said unknown variables. In our application, we select disjoint groups of $G \leq N$ independent entries in $\boldsymbol{w}$; the larger the group size the more dependency structure is retained and, in general, the more accurate the achieved approximation will be. On the other hand, by selecting groups with dimension $G << N$, we are able to significantly reduce the computational complexity of the resulting SBL algorithm. Our goal is, thus, to investigate if small group sizes can be selected without reducing the recovery performance of the SBL algorithm. We test our proposed algorithms by applying them to the generic signal model (1) and for the estimation of the wireless channel in an orthogonal frequency-division multiplexing (OFDM) receiver. Our reported numerical results show that a significant reduction in complexity can be achieved with no significant penalization in performance with respect to both mean-squared error (MSE) of the channel estimates and bit-error-rate (BER).

II. GMF For SBL

In this section we present the GMF-based SBL algorithms. The first step is to state the joint pdf for the signal model (1). Based on this probabilistic model, we derive the update rules for GMF inference. The approach presented is general in the sense that it can be used with a large variety of prior models. In the end of the section we show how, by appropriately setting the parameters of the chosen prior model, we can obtain different low complexity versions of a variety of SBL algorithms.

A. Probabilistic Model

We make use of a two-layer hierarchical representation of the prior $p(\boldsymbol{w})$ involving a conditional prior $p(\boldsymbol{w}|\boldsymbol{\gamma})$ and a hyperprior $p(\boldsymbol{\gamma})$. The joint pdf for the signal model (1)

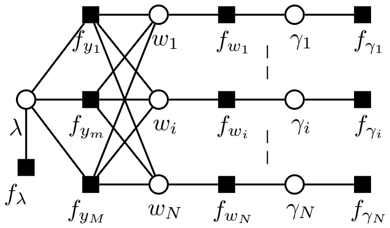


Fig. 1. Factor graph representation of the joint pdf (2); $f_{y_m} \triangleq p(y_m|\boldsymbol{w}, \lambda)$, $f_{w_i} \triangleq p(w_i|\gamma_i)$, and $f_{\gamma_i} \triangleq p(\gamma_i)$.

augmented with this prior model then reads:

$$p(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{\gamma}, \lambda) = p(\lambda) \prod_{m=1}^{M} p(y_m|\boldsymbol{w}, \lambda) \prod_{i=1}^{N} p(w_i|\gamma_i)p(\gamma_i). \quad (2)$$

The hierarchical representation of $p(\boldsymbol{w})$ effectively circumvents possible intractable computation of the posterior $p(\boldsymbol{w}|\boldsymbol{y})$ as we are free to select "simple" pdfs for $p(w_i|\gamma_i)$ and $p(\gamma_i)$. We follow our approach in [7] and consider the hierarchical representation of the Bessel K pdf by letting $p(w_i|\gamma_i) = \mathrm{N}(w_i|0, \gamma_i)$ and $p(\gamma_i) = \mathrm{Ga}(\gamma_i|\epsilon, \eta)$.[1] For the noise precision λ, we select the noninformative Jeffreys prior, $p(\lambda) \propto 1/\lambda$. Finally, due to (1), $p(y_m|\boldsymbol{w}, \lambda) = \mathrm{N}(y_m|\sum_i \phi_{mi} w_i, \lambda^{-1})$.

B. GMF Approximation

Let $\boldsymbol{\theta} = \{\boldsymbol{w}, \boldsymbol{\gamma}, \lambda\}$ be the set of unknown parameters to be estimated. The mean field (MF) approximation refers to variational methods that attempt to approximate the true density $p(\boldsymbol{\theta}|\boldsymbol{y})$ with an auxiliary pdf $b(\boldsymbol{\theta})$ by minimizing the Kullback-Leibler (KL) divergence $\mathrm{KL}(b(\boldsymbol{\theta})\|p(\boldsymbol{\theta}|\boldsymbol{y}))$, see e.g., [11]. We are free to select a structure of $b(\boldsymbol{\theta})$ that allows for a simple and computationally efficient update of $b(\boldsymbol{\theta})$. As we will see, the key to achieve this is to define disjoint groups of entries in $\boldsymbol{w}$. We define our auxiliary pdf as a structured factorization [8]–[10] according to

$$b(\boldsymbol{\theta}) = \prod_k b(\boldsymbol{\theta}_k) = b(\lambda) \prod_{i=1}^{N} b(\gamma_i) \prod_{q=1}^{Q} b(\boldsymbol{w}_q) \quad (3)$$

with the vector $\boldsymbol{w}_q \triangleq [w_i|i \in \{(q-1)G+1 : qG\}]^{\mathrm{T}}$, $q \in \{1 : Q\}$, representing disjoint groups of G contiguous entries in $\boldsymbol{w}$ and $N = QG$. From (3), we obtain the naive MF approximation – i.e., with $b(\boldsymbol{\theta})$ being a fully factorized function – by setting $G = 1$ and having, thus, $Q = N$ groups of a single entry. Conversely, the fully structured MF approximation is obtained with $G = N$ and, thus, $Q = 1$. Notice that, due to the construction of the prior model for $p(\boldsymbol{w})$, the inferred form of $b(\boldsymbol{\gamma})$, which we detail later in this section, factorizes according to $b(\boldsymbol{\gamma}) = \prod_i b(\gamma_i)$, regardless of whether this factorization is explicitly imposed in (3) or not. However, this is not the case for $b(\boldsymbol{w})$ because of the factors $p(y_m|\boldsymbol{w}, \lambda)$, $m = 1, \ldots, M$. The factor graph depicted in Fig. 1 visualizes the statistical dependency of the variables in the probabilistic model (2).

Our goal is to analyze the effect of different factorizations of (3) on the accuracy and computational complexity of different SBL algorithms. Generally speaking, one would expect the accuracy of the estimates to degrade with finer factorizations (decreasing G), as the space of functions over which the KL divergence is minimized becomes more restricted; on the other hand, finer factorizations often yield algorithms with lower computational complexity than the algorithms based on coarser factorizations.

The update rule for the kth factor of the GMF approximation (3) can be written in the simple form [12]

$$b(\boldsymbol{\theta}_k) \propto \exp\left(\langle \log p(\boldsymbol{y}, \boldsymbol{\theta})\rangle_{\prod_{l \neq k} b(\boldsymbol{\theta}_l)}\right), \quad (4)$$

where the expression $\langle f(\boldsymbol{x})\rangle_{p(\boldsymbol{x})}$ denotes the expectation of a function $f(\boldsymbol{x})$ with respect to a density $p(\boldsymbol{x})$. After an initialization procedure, each algorithmic iteration consists of sequentially computing all individual factors $b(\boldsymbol{\theta}_k)$ of $b(\boldsymbol{\theta})$.

From (4), the factor $b(\boldsymbol{w}_q)$ is a normal pdf with mean $\boldsymbol{\mu}_q$ and covariance $\boldsymbol{\Sigma}_q$ given by

$$\boldsymbol{\mu}_q = \boldsymbol{\Sigma}_q \langle \lambda \rangle_{b(\lambda)} \boldsymbol{\Phi}_q^{\mathrm{H}} (\boldsymbol{y} - \sum_{q' \neq q} \boldsymbol{\Phi}_{q'} \boldsymbol{\mu}_{q'}), \quad (5)$$

$$\boldsymbol{\Sigma}_q = \left(\langle \lambda \rangle_{b(\lambda)} \boldsymbol{\Phi}_q^{\mathrm{H}} \boldsymbol{\Phi}_q + \langle \boldsymbol{\Gamma}_q^{-1} \rangle_{b(\boldsymbol{\gamma})}\right)^{-1}, \quad (6)$$

where $\boldsymbol{\Gamma}_q = \mathrm{diag}(\boldsymbol{\gamma}_q)$ with $\boldsymbol{\gamma}_q$ defined analogously to $\boldsymbol{w}_q$ and $\boldsymbol{\Phi}_q \triangleq [\boldsymbol{\phi}_i|i \in \{(q-1)G+1 : qG\}]$. We define $\boldsymbol{\mu} \triangleq [\boldsymbol{\mu}_1^{\mathrm{T}}, \ldots, \boldsymbol{\mu}_Q^{\mathrm{T}}]^{\mathrm{T}}$ and $\boldsymbol{\Sigma}$ as the block diagonal matrix $\boldsymbol{\Sigma} \triangleq \mathrm{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_Q)$. From $b(\boldsymbol{w}) = \prod_q b(\boldsymbol{w}_q)$, we produce a point estimate of $\boldsymbol{w}$ as $\hat{\boldsymbol{w}} = \boldsymbol{\mu}$.

The computational complexity of the GMF-based SBL algorithms is determined by the updates (5) and (6). In big-O notation the complexity is $\max\{O(\hat{K}G^2), O(\hat{K}^2)\}$ per algorithmic iteration, where $\hat{K}$ denotes the nonzero entries in $\boldsymbol{\mu}$. Naturally, the algorithm can remove a vector $\boldsymbol{\phi}_i$ once the corresponding $\langle \gamma_i^{-1} \rangle_{b(\gamma_i)}$ becomes large enough [4], which drastically reduces the computational complexity of the update (6). However, in the first iterations $\hat{K} = N$. This emphasizes the importance of grouping entries in $\boldsymbol{w}$ in order to reduce the computational complexity of the initial iterations of the algorithm.

The auxiliary function $b(\lambda)$ can be shown to be a gamma pdf with mean

$$\langle \lambda \rangle_{b(\lambda)} = \frac{M}{\langle \|\boldsymbol{y} - \boldsymbol{\Phi}\boldsymbol{w}\|_2^2 \rangle_{\prod_q b(\boldsymbol{w}_q)}}. \quad (7)$$

Note that the update of λ is often neglected in other inference schemes, such as belief propagation, since a simple, tractable expression cannot be achieved.

In the following, we particularize our GMF algorithm by specifying the parameters of the prior model in (2) (corresponding to the selection of the parameters ϵ and η in $p(\gamma_i)$). We select the parameters appropriately to obtain low complexity versions of different SBL algorithms. Selecting a

[1]For a real (complex) random vector $\boldsymbol{x}$, $\mathrm{N}(\boldsymbol{x}|\boldsymbol{a}, \boldsymbol{B})$ denotes the real (complex) multivariate normal pdf with mean $\boldsymbol{a}$ and covariance matrix $\boldsymbol{B}$. Similarly, $\mathrm{Ga}(x|a, b) = \frac{b^a}{\Gamma(a)} x^{a-1} \exp(-bx)$ is a Gamma density.

group size of $G = N$ for $b(\boldsymbol{w})$ leads to the original algorithms found in [4], [7], [13]. These inference methods only differ from each other in the update of $b(\boldsymbol{\gamma}) = \prod_i b(\gamma_i)$. Observe that the computation of $\boldsymbol{\Sigma}$ requires evaluating $\langle \gamma_i^{-1} \rangle_{b(\gamma_i)}$ for all $i = 1, \ldots, N$. We review these updates in the following.

GMF-RVM: The RVM algorithm [4] ($G = N$) results from selecting the noninformative Jeffreys prior for each γ_i [12]. By selecting $\epsilon = \eta = 0$, $p(\gamma_i)$ reduces to this improper prior. In this way, $b(\boldsymbol{\gamma})$ becomes a product of N inverse gamma pdfs. The update of $\langle \gamma_i^{-1} \rangle_{b(\gamma_i)}$ then follows as

$$\langle \gamma_i^{-1} \rangle_{b(\gamma_i)} = \frac{1}{\Sigma_{ii} + |\mu_i|^2}, \quad i = 1, \ldots, N. \tag{8}$$

GMF-BPDN: Basis pursuit denoising (BPDN) [1], [2] refers to the solution of

$$\underset{\boldsymbol{w}}{\operatorname{argmin}} \left\{ \rho \|\boldsymbol{y} - \boldsymbol{\Phi} \boldsymbol{w}\|_2^2 + \kappa \|\boldsymbol{w}\|_1 \right\}, \tag{9}$$

where κ is some positive regularization constant. We have introduced the parameter ρ to distinguish between two cases: $\rho = 1/2$ when $\boldsymbol{y}, \boldsymbol{\Phi}, \boldsymbol{w}, \boldsymbol{n}$ in (1) are all real and $\rho = 1$ when they are complex. We can solve the optimization problem (9) using iterative Bayesian inference by selecting the prior model of $p(\boldsymbol{w})$ as a hierarchical representation of N Laplace pdfs and formulating an algorithm based on the expectation-maximization algorithm with complete data $\{\boldsymbol{y}, \boldsymbol{\gamma}\}$. The former corresponds to setting $\epsilon = \rho + 1/2$ in (2) [7], while the latter can be achieved by constraining the approximating factor $b(\boldsymbol{w})$ in the GMF framework to represent the point estimate $\hat{\boldsymbol{w}} = \boldsymbol{\mu}$, i.e., setting $b(\boldsymbol{w}) = \delta(\boldsymbol{w} - \hat{\boldsymbol{w}})$ with $\delta(\cdot)$ denoting the Dirac delta function [14]. By doing so, we obtain

$$\langle \gamma_i^{-1} \rangle_{b(\gamma_i)} = \frac{\sqrt{\eta/\rho}}{|\mu_i|}, \quad i = 1, \ldots, N. \tag{10}$$

Selecting $G = N$ and $\rho = 1/2$ yields the algorithm proposed in [13].

GMF-BesselK: In this SBL algorithm, proposed in [7] ($G = N$), we solve for $b(\boldsymbol{\gamma})$ without setting the parameters ϵ and η of $p(\gamma_i)$ a priori. This makes $b(\boldsymbol{\gamma})$ a product of N generalized inverse Gaussian (GIG) pdfs. The moments of a GIG pdf can be computed in closed form that involves the modified Bessel function of the second kind. As we target low complexity algorithms, we compute the mode instead by restricting $b(\boldsymbol{\gamma}) = \delta(\boldsymbol{\gamma} - \hat{\boldsymbol{\gamma}})$:

$$\langle \gamma_i^{-1} \rangle_{b(\gamma_i)} = \frac{(\rho + 1 - \epsilon) + \sqrt{\Delta_i}}{2\rho(\Sigma_{ii} + |\mu_i|^2)}, \tag{11}$$

with $\Delta_i = (\rho + 1 - \epsilon)^2 + 4\rho\eta(\Sigma_{ii} + |\mu_i|^2)$ and ρ defined as in (9).

III. Numerical Results

We perform Monte Carlo simulations to investigate the impact of different factorizations of $b(\boldsymbol{w}) = \prod_q b(\boldsymbol{w}_q)$ on the performance of the proposed GMF-based SBL algorithms described in Section II. We first consider a generic signal model (1) commonly used in sparse signal representation. We then apply the GMF-based algorithms for the estimation of the wireless channel in an OFDM system.

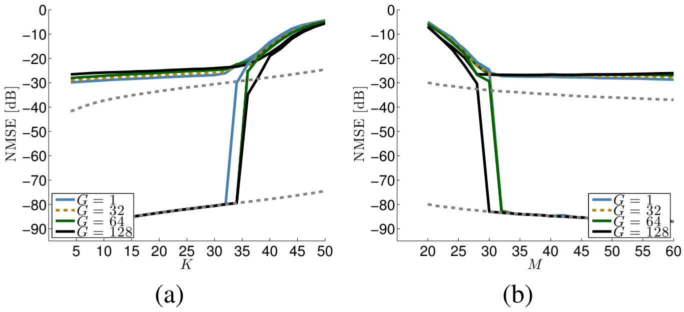


Fig. 2. Comparison of the NMSE achieved by GMF-RVM with different group sizes G and SNR as a parameter. We have $N = 128$, (a) $M = 64$, and (b) $K = 10$. The SNR values: 30 dB and 80 dB.

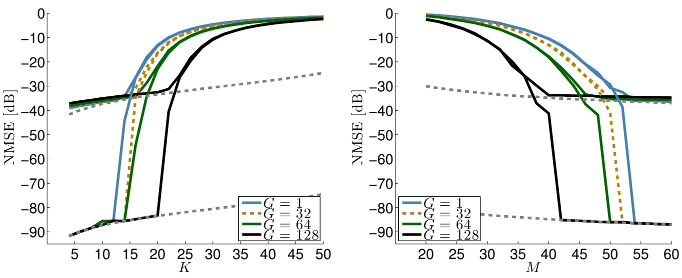
In all setups, the GMF-based SBL algorithms are initialized with $\langle \lambda \rangle_{b(\lambda)} = 1/\mathbb{V}\mathrm{ar}(\boldsymbol{y})$ and $\langle \gamma_i^{-1} \rangle_{b(\gamma_i)} = 1$, $i = 1, \ldots, N$. As the iterations proceed, an entry μ_i is set to zero when $\langle \gamma_i^{-1} \rangle_{b(\gamma_i)}$ exceeds a fixed threshold set at 10^6, and the corresponding vector $\boldsymbol{\phi}_i$ is removed from the dictionary matrix $\boldsymbol{\Phi}$. Once the initialization is completed, the algorithm sequentially updates the auxiliary pdfs $b(\boldsymbol{w}_q)$, $q = 1, \ldots, Q$, $b(\boldsymbol{\gamma})$, and $b(\lambda)$ until $\|\boldsymbol{\mu}^+ - \boldsymbol{\mu}\|_\infty \leq 10^{-8}$, where $\boldsymbol{\mu}^+$ and $\boldsymbol{\mu}$ denote the mean of $b(\boldsymbol{w})$ for two consecutive iterations.

A. Sparse Signal Representation

For the signal model (1), the entries in $\boldsymbol{\Phi}$ are independent and identically distributed (iid) zero-mean complex normal with variance M^{-1}. Similarly, the K nonzero entries in $\boldsymbol{w}$ are iid zero-mean complex normal with variance one, with their indices being uniformly drawn from the range $\{1 : N\}$. As a reference, we include the performance of the oracle estimator that "knows" the indices of the K nonzero entries in $\boldsymbol{w}$ and computes a least-squares estimate of these entries (grey dashed curve in the subsequent figures). All reported results are computed based on a total of 1000 Monte Carlo runs.

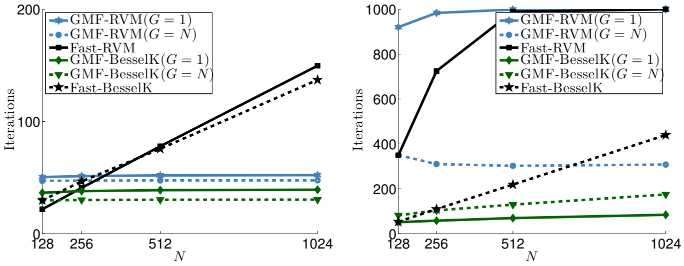
We will see that the impact of the group size G on the estimation performance strongly depends on the prior model (selection of ϵ and η) used to derive the corresponding GMF-based SBL algorithm. To demonstrate this, we evaluate the performance for different signal-to-noise-ratios (SNRs), number of observations M, and number of nonzero entries K.

Fig. 2 compares the normalized mean-squared error (NMSE), $\mathrm{NMSE} \triangleq \langle \|\boldsymbol{w} - \hat{\boldsymbol{w}}\|_2^2 \rangle / \langle \|\boldsymbol{w}\|_2^2 \rangle$, achieved by GMF-RVM($G$) with different group sizes $G \in \{1, N/4, N/2, N\}$ versus (a) K and (b) M. The dimension of $\boldsymbol{w}$ is $N = 128$. In (a), we have $M = 64$ and in (b) $K = 10$. The SNR is set to 30 dB and 80 dB. Interestingly, the conditions with respect to K and M under which the signal $\boldsymbol{w}$ can be recovered seem to be independent of the SNR and no significant difference in performance is observed between the chosen group sizes. Thus, GMF-RVM($G = 1$) exhibits a performance similar to that of the "traditional" RVM ($G = N$) [4] but with a reduction in complexity from $O(\hat{K}^3)$ to $O(\hat{K}^2)$.



(a) (b)

Fig. 3. Comparison of the NMSE achieved by GMF-BesselK with different group sizes G and SNR as a parameter. We have $N = 128$, (a) $M = 64$, and (b) $K = 10$. The SNR values: 30 dB and 80 dB.



(a) SNR at 80 dB. (b) SNR at 30 dB.

Fig. 4. Comparison of the convergence rate achieved by GMF-RVM and GMF-BesselK with different group sizes G. We have $N \in \{128, 256, 512, 1024\}$, $M = N/2$, and $K = \lceil N/10 \rceil$.

We perform the same experiment for GMF-BesselK with $\epsilon = 1/2$ and $\eta = 1$ in Fig. 3. Again we observe the same threshold-like behavior in the NMSE curves that is independent of the SNR, but a performance loss is incurred when G is reduced. However, if the signal is sparse enough and the number of measurements M, is sufficiently large, we can significantly reduce G with no penalization in performance.

The analogous simulations were also conducted for GMF-BPDN with similar conclusions made as for GMF-RVM. For the sake of brevity, we have omitted the results.

Finally, it is important to check whether the reduction in complexity per algorithm iteration comes at the expense of a higher iteration count before convergence is reached. For this comparison, we also include Fast-RVM [5][2] and Fast-BesselK [7] (with $\epsilon = 1/2$ and $\eta = 1$). These greedy methods have a complexity of $O(MN\hat{K})$ per algorithmic iteration. The stopping criterion used is identical to that of the GMF algorithms. Fig. 4 shows the result as a function of the problem size: $N \in \{128, 256, 512, 1024\}$, $M = N/2$, and $K = \lceil N/10 \rceil$. Several remarks are worth noting. First, by construction, the iteration count for greedy algorithms inherently depends on K. In high SNR regime (Fig. 4(a)), we observe that the GMF-based algorithms do not suffer from this. For $G = 1$ the count is of the same order as that of the high complexity algorithms with $G = N$. Second, by comparing Figs. 4(a)-4(b), we observe that the iteration count is heavily affected by the SNR. This is especially true for the GMF-RVM algorithms: GMF-RVM($G = 1$) experiences a slow convergence rate.[3] On the other hand, GMF-BesselK($G = 1$) achieves the lowest iteration count of all algorithms. This indicates that the rate of convergence of a particular algorithm is dominated by the prior model used to derive it rather than the choice of a specific group size G.

B. Sparse Channel Estimation in an OFDM Receiver

We next apply the GMF-based algorithms to the problem of pilot-assisted channel estimation in OFDM systems. We only consider GMF-BesselK for these investigations as our previously reported numerical results show that GMF-BesselK clearly outperforms the other GMF-based algorithms with respect to speed of convergence.

A single-input–single-output OFDM system is considered with a cyclic prefix (CP) inserted to eliminate inter-symbol interference. The channel response is assumed static during the transmission of each OFDM block. The received baseband signal $\boldsymbol{r} \in \mathbb{C}^{M_u}$ is given by

$$\boldsymbol{r} = \boldsymbol{X}\boldsymbol{h} + \boldsymbol{n}. \tag{12}$$

Here, $\boldsymbol{X} = \operatorname{diag}(\boldsymbol{x})$ contains the complex-modulated symbols $\boldsymbol{x} \in \mathbb{C}^{M_u}$ and the entries in $\boldsymbol{n} \in \mathbb{C}^{M_u}$ are iid zero-mean complex normal with variance λ^{-1}. The vector $\boldsymbol{h}$ contains the samples of the channel frequency response at all M_u subcarriers. Let the set $\mathcal{P} \subseteq \{1, \dots, M_u\}$ contain the indices of the subcarriers reserved for pilot transmission. The $M \triangleq |\mathcal{P}| < M_u$ pilot observations used for estimating $\boldsymbol{h}$ are then

$$\boldsymbol{y} \triangleq (\boldsymbol{X}_{\mathcal{P}})^{-1}\boldsymbol{r}_{\mathcal{P}} = \boldsymbol{h}_{\mathcal{P}} + \tilde{\boldsymbol{n}}, \tag{13}$$

where $\boldsymbol{r}_{\mathcal{P}} = [r_m : m \in \mathcal{P}]^{\mathrm{T}}$ and $\boldsymbol{h}_{\mathcal{P}} = [h_m : m \in \mathcal{P}]^{\mathrm{T}}$. The statistics of the noise term $\tilde{\boldsymbol{n}} \triangleq (\boldsymbol{X}_{\mathcal{P}})^{-1}\boldsymbol{n}_{\mathcal{P}}$ remain unchanged as the pilot symbols hold unit power.

In order to apply sparse methods for estimating $\boldsymbol{h}$ in (12) we must assume some basis in which $\boldsymbol{h}$ is sparse or approximately so and then recast the OFDM pilot observation model (13) into the form of (1). Hence, a dictionary $\boldsymbol{\Phi}$ for $\boldsymbol{h}$ must be constructed. For doing so, we follow the common assumption that the wireless multipath channel is sparse in the delay domain and consider a frequency-selective wireless channel with impulse response modeled as a sum of specular multipath components:

$$g(\tau) = \sum_{k=1}^{K} \beta_k \delta\left(\tau - \tau_k\right). \tag{14}$$

The entries of the vectors $\boldsymbol{\beta} = [\beta_1, \dots \beta_K]$ and $\boldsymbol{\tau} = [\tau_1, \dots, \tau_K]$ are respectively the complex weights and the delays of the K multipath components. Given (14), $\boldsymbol{h}$ can be written as $\boldsymbol{h} = \boldsymbol{\Phi}(\boldsymbol{\tau})\boldsymbol{\beta}$ with $\boldsymbol{\Phi}(\boldsymbol{\tau})_{m,k} = \exp\left(-j2\pi f_m \tau_k\right)$

[2]We experienced that Fast-RVM overestimates the noise precision which produces non-sparse estimates. As a result, we let $\hat{\lambda} = \lambda$.

[3]The algorithms terminate if a maximum of 1000 iterations are reached.

TABLE I
PARAMETER SETTINGS FOR THE SIMULATIONS.

Sampling time, T_s	32.55 ns
CP length	4.69 μs / 144 T_s
Subcarrier spacing	15 kHz
Pilot pattern	Evenly spaced, QPSK
Modulation	QPSK ($M_d = 2$)
Subcarriers, M_u	1200
OFDM symbols	1
Information bits	1091
Channel interleaver	Random
Convolutional code	$(133, 171, 165)_8$
Decoder	BCJR algorithm [15]

and f_m denoting the frequency of the mth subcarrier, $m = 1, \ldots, M_u$. However, as the delays are unknown, $\boldsymbol{\Phi}(\boldsymbol{\tau})$ is unknown to the algorithms. We therefore construct a dictionary according to $\boldsymbol{\Phi}(\boldsymbol{\tau}_d)_{m,i} = \exp(-j2\pi f_m \tau_{d_i})$, $i = 1, \ldots, N$, where the entries in $\boldsymbol{\tau}_d \in \mathbb{R}_+^N$ are delay samples uniformly-spaced in the interval $[0, \tau_{\max}]$:

$$\boldsymbol{\tau}_d = \left[0, \frac{T_s}{\zeta}, \frac{2T_s}{\zeta}, \ldots, \tau_{\max}\right]^{\mathrm{T}} \tag{15}$$

with $\zeta > 0$ such that $N = \zeta\tau_{\max}/T_s + 1$ is an integer. The symbols $\tau_{\max}$ and T_s denote respectively the maximum excess delay of the channel and the sampling time.

We can now apply sparse representation methods to the approximate signal model

$$\boldsymbol{y} = \boldsymbol{h}_{\mathcal{P}} + \tilde{\boldsymbol{n}} \approx \boldsymbol{\Phi}_{\mathcal{P}}(\boldsymbol{\tau}_d)\boldsymbol{w} + \tilde{\boldsymbol{n}} \tag{16}$$

with $\boldsymbol{\Phi}_{\mathcal{P}}(\boldsymbol{\tau}_d)$ containing the rows of $\boldsymbol{\Phi}(\boldsymbol{\tau}_d)$ corresponding to the indices in $\mathcal{P}$. The final estimate of $\boldsymbol{h}$ is then $\hat{\boldsymbol{h}} \triangleq \boldsymbol{\Phi}(\boldsymbol{\tau}_d)\hat{\boldsymbol{w}}$. Hence, we seek to accurately represent $\boldsymbol{h}$ in (12) using the sparse approximation $\hat{\boldsymbol{h}}$.

We consider an OFDM transmission scenario inspired by the 3GPP LTE standard [16] with the settings specified in Table I. In all conducted investigations we fix the spectral efficiency to $M_d(M_u - M)R/M_u = 0.92$ information bits per subcarrier, which corresponds to a rate $R = 1/2$ code obtained through puncturing. Unless otherwise specified, we set the number of rows in $\boldsymbol{\Phi}_{\mathcal{P}}(\boldsymbol{\tau}_d)$ to $M = 100$ (pilot subcarriers) and the number of columns to $N = 200$, which corresponds to a delay resolution of $T_s/\zeta = 0.72\ T_s$ (≈ 23.4 ns) and $\tau_{\max} = 144\ T_s$ (the CP length).

GMF-BesselK is tested with three group sizes $G \in \{1, 10, N\}$. For comparison we include two non-Bayesian methods, BPDN and orthogonal matching pursuit (OMP), see e.g., [3]. We also conducted experiments with Fast-BesselK but we obtained similar performance as GMF-BesselK($G = N$), so these results are not shown. For BPDN, we use the sparse reconstruction by separable approximation (SpaRSA) algorithm [17]. The required regularization parameter is chosen as $5\sqrt{\log(N)/\lambda}$. For OMP we set the number of multipath components to search for to 20. These settings empirically led to satisfactory results. The commonly employed robustly designed Wiener filter (RWF) [18] for OFDM channel estimation is also included as a reference.

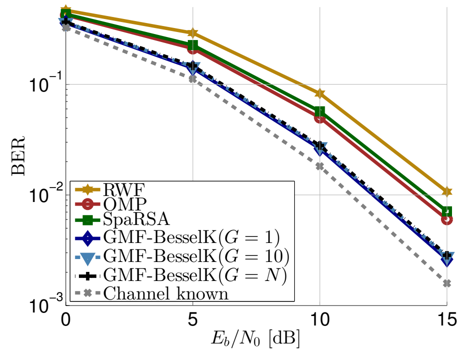


Fig. 5. Comparison of the BER of the OFDM receiver incorporating the different algorithms using $M = 100$ pilot symbols. The channel parameters are $1/V = 300$ ns, $1/v = 5$ ns, $U = 60$ ns, and $u = 20$ ns.

The above channel estimators are embedded in an OFDM receiver that decodes the transmitted information bits using a BCJR algorithm. The performance of the channel estimators (in terms of MSE) and of the corresponding receiver (in terms of BER) are assessed by means of Monte Carlo simulations. Channel impulse responses are generated independently using the model proposed by Saleh and Valenzuela [19] for indoor environments:

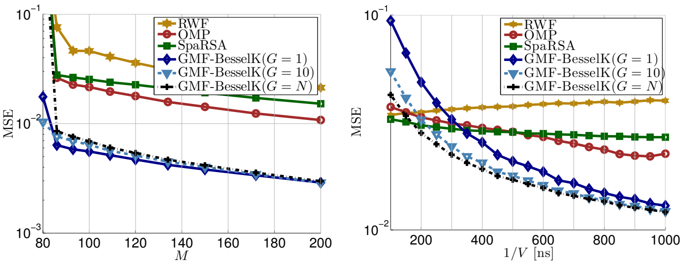
$$g(\tau) = \sum_{l=0}^{\infty}\sum_{k=0}^{\infty} \beta_{k,l}\delta\left(\tau - (T_l + \tau_{k,l})\right). \tag{17}$$

Here, $\{T_l\}_l$ (cluster delays) and $\{\tau_{k,l}\}_k$ (within cluster delays) are both homogeneous Poisson processes with rate parameter V and v respectively. Conditioned on $\{T_l\}_l$ and $\{\tau_{k,l}\}_k$, $\{\beta_{k,l}\}_{k,l}$ are independent zero-mean complex normal distributed with variance

$$\sigma^2(T_l, \tau_{k,l}) = Q\exp(-T_l/U)\exp(-\tau_{k,l}/u). \tag{18}$$

We compute Q such that $\langle \sum_l \sum_k |\beta_{k,l}|^2 \rangle = 1$. It is important to stress that the specular channel model (14) has inspired the design of the dictionary matrix, while the Saleh and Valenzuela model (17) is used in the performance assessment.

We follow [19] and select the channel parameters according to $1/V = 300$ ns, $1/v = 5$ ns, $U = 60$ ns, and $u = 20$ ns. From this, we have on average a spacing of 300 ns between cluster delays and 5 ns between within cluster delays. The parameters U and u ensures that the power of the multipath components exhibits a fast decay relatively to the CP length typically encountered in an indoor scenario. The BER performance is depicted in Fig. 5. Clearly, the GMF-BesselK algorithms lead to better performance than the other channel estimators. At 1 % BER, the gain is 2 dB over OMP and SpaRSA, and 3 dB over RWF. No performance drop is observed for GMF-BesselK when decreasing the group size G as the GMF-BesselK algorithms reconstruct $\boldsymbol{h}$ properly from only approximately 5-10 column vectors in $\boldsymbol{\Phi}(\boldsymbol{\tau}_d)$ across SNR (results not shown). We also evaluated the MSE performance of the channel estimators, defined as $\mathrm{MSE} \triangleq \langle\|\boldsymbol{h} - \hat{\boldsymbol{h}}\|_2^2\rangle/M_u$,



(a) (b)

Fig. 6. Comparison of the MSE achieved by the different algorithms versus (a) number of pilot symbols M and (b) cluster rate $1/V$. In (a) the channel parameters are $1/V = 300$ ns, $1/v = 5$ ns, $U = 60$ ns, and $u = 20$ ns. In (b) $M = 100$, and we have $1/v = 5$ ns, $U = 900$ ns, and $u = 20$ ns.

versus the number of pilots M. The results depicted in Fig. 6(a) show the superior performance of GMF-BesselK and illustrate that, even though the model (17) is not sparse, it is compressible such that a proper sparse approximation can be achieved by the estimators.

Based on the above results, we next compare the algorithms versus the number of cluster components. To ensure a longer maximum excess delay, we set $U = 900$ ns. The parameters v and u are selected as before. In Fig. 6(b) we show the MSE versus the cluster rate parameter $1/V = 1 : 1000$ ns.[4] When $1/V \geq 800$ ns, the performance of GMF-BesselK($G = 1$) is on par with GMF-BesselK($G = N$), but for $1/V \leq 800$ the performance of GMF-BesselK($G = 1$) drops as compared to GMF-BesselK($G = N$). However, this break in performance is mitigated using only a group size of $G = 10$. This setting allows for a significant decrease in computational complexity as compared to using $G = N$.

IV. Conclusion

We have proposed the use of generalized mean field (GMF) inference for low complexity implementations of a wide range of sparse Bayesian learning (SBL) algorithms. More specifically, we use the GMF approach to approximate the posterior probability density function (pdf) of the sparse weight vector with a simpler auxiliary pdf, which factorizes over disjoint groups of entries in this vector. The approach presented in this paper yields simple and low complexity expressions for the parameter updates, is valid for the estimation of real- and complex-valued signals, and is general in the sense that it can be applied to many SBL algorithms. At the expense of less dependency structure in the auxiliary pdf, the resulting GMF-based SBL algorithms lead to a significant reduction in the computational complexity as compared to their original counterparts.

The numerical assessment shows that the complexity reduction can be achieved with no significant performance degradation. The investigations were conducted for two scenarios: application to a generic compressive sensing signal model and estimation of the wireless channel in an orthogonal frequency-division multiplexing receiver. They revealed that the impact of the factorizations of the auxiliary pdf on the algorithms' performance highly depends on the underlying prior model of the sparse weight vector. For the latter scenario, the numerical results show that the proposed algorithms outperform state-of-the-art non-Bayesian inference algorithms for sparse channel estimation.

[4]For OMP we decreased the number of components to search for as $1/V$ increased; specifically, we selected: $\{50, 48, \ldots\}$.

Acknowledgment

This work was supported in part by the 4GMCT cooperative research project, funded by Intel Mobile Communications, Agilent Technologies, Aalborg University and the Danish National Advanced Technology Foundation.

References

[1] R. Tibshirani, "Regression shrinkage and selection via the LASSO," *J. R. Statist. Soc.*, vol. 58, pp. 267–288, 1994.
[2] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. on Scientific Computing*, vol. 20, pp. 33–61, 1998.
[3] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. on Inf. Theory*, vol. 50, pp. 2231–2242, Oct. 2004.
[4] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. of Machine Learning Research*, vol. 1, pp. 211–244, 2001.
[5] M. Tipping and A. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proc. 9th Int. Workshop on Artificial Intelligence and Statistics*, 2003.
[6] S. Babacan, R. Molina, and A. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. on Image Proc.*, vol. 19, no. 1, pp. 53–63, 2010.
[7] N. Pedersen, D. Shutin, C. Manchón, and B. Fleury, "Sparse estimation using Bayesian hierarchical prior modeling for real and complex models," *in preparation*, 2013, arXiv:1108.4324.
[8] E. Xing, M. Jordan, and S. Russell., "A generalized mean field algorithm for variational inference in exponential families," *in Uncertainty in Artificial Intelligence*, vol. 19, 2003.
[9] C. Bishop and J. Winn, "Structured variational distributions in VIBES," in *Artificial Intelligence and Statistics*, 2003. Society for Artificial Intelligence and Statistics.
[10] J. Dauwels, "On variational message passing on factor graphs," *in IEEE Int. Sym. on Inform. Theory (ISIT'07)*, pp. 2546–2550, 2007.
[11] J. Winn and C. Bishop, "Variational message passing," *J. of Machine Learning Research*, vol. 6, pp. 661–694, 2005.
[12] C. Bishop and M. Tipping, "Variational relevance vector machines," *in Proc. of the 16th Conference on Uncertainty in Artificial Intelligence*, pp. 46–53, 2000.
[13] M. Figueiredo, "Adaptive sparseness for supervised learning," *IEEE Trans. on Pattern Analysis and Machine Intel.*, vol. 25, no. 9, pp. 1150–1159, 2003.
[14] E. Riegler, G. Kirkelund, C. Manchón, M.-A. Badiu, and B. Fleury, "Merging belief propagation and the mean field approximation: A free energy approach," *IEEE Trans. on Inf. Theory*, vol. 59, no. 1, pp. 588–602, 2013.
[15] L. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. on Inf. Theory*, vol. 20, no. 2, pp. 284–287, 1974.
[16] 3rd Generation Partnership Project (3GPP) Technical Specification, "Evolved universal terrestrial radio access (e-utra); base station (bs) radio transmission and reception," TS 36.104 V8.4.0, Tech. Rep., 2008.
[17] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. on Sig. Proc.*, vol. 57, no. 7, pp. 2479–2493, 2009.
[18] O. Edfors, M. Sandell, J.-J. van de Beek, S. K. Wilson, and P. O. Börjesson, "OFDM channel estimation by singular value decomposition," *IEEE Trans. on Communications*, vol. 46, no. 7, pp. 931–939, 1998.
[19] A. A. Saleh and R. Valenzuela, "A statistical model for indoor multipath propagation," *IEEE Journal on Selected Areas in Communications*, vol. 5, no. 2, pp. 128–137, 1987.