Structured and soft ! Boltzmann machine and mean-field approximation for structured sparse representations

Angélique Drémeau [a,b]
[a]Fondation Pierre-Gilles De Gennes pour la Recherche
29 rue d'Ulm, F-75005 Paris, France
Email: angelique.dremeau@espci.fr

Laurent Daudet [b]
[b]Institut Langevin, ESPCI ParisTech, Univ. Paris Diderot, CNRS
UMR 7587, F-75005 Paris, France
Email: laurent.daudet@espci.fr

***Abstract*—We address the problem of *structured* sparse representation within a Bayesian framework. In particular, we consider a mean-field approximation for the estimation of the dependencies between atoms using a Boltzmann machine. This algorithm is shown to outperform the reference algorithm [1] with regard to their success criterion.**

***Index Terms*—Structured sparse representations, Boltzmann machine, mean-field approximation.**

I. Introduction

Recent contributions have emphasized the interest of considering *structures* between atoms selected in sparse representations (SR), for a wide range of dictionaries and classes of signals. This problem can be set into a Bayesian framework, e.g. Cevher *et al.* [2] and Faktor *et al.* [1]. Both use Boltzmann machines to model the dependencies between atoms, but differ in the prior model on the SR coefficients. In this paper, we consider a similar model as in [1].

Our observation model is $\mathbf{y} = \sum_{i=1}^{M} s_i\, x_i\, \mathbf{d}_i + \mathbf{n}$, where $\mathbf{s} \in \{0,1\}^M$ is the SR support, $\mathbf{n} \sim \mathcal{N}(0, \sigma_n^2 \mathbf{I}_N)$ and $\mathbf{I}_N$ the identity matrix. We suppose that $\forall i$, $p(x_i) = \mathcal{N}(0, \sigma_{x_i}^2)$ and $\mathbf{s}$ is distributed according to a Boltzmann machine of parameters $\mathbf{b}$ and $\mathbf{W}$[1]:

$$p(\mathbf{s}) \propto \exp(2\mathbf{b}^T\mathbf{s} + 2\mathbf{s}^T\mathbf{W}\mathbf{s} - 2\mathbb{1}_M^T\mathbf{W}\mathbf{s}), \tag{1}$$

where $\mathbb{1}_M = [1, \ldots, 1]$ of length M.

II. Structured Soft Bayesian Pursuit Algorithm

Based on this model, we consider here the following *marginalized* maximum a posteriori (MAP) estimation problem:

$$\hat{\mathbf{s}} = \underset{\mathbf{s}\in\{0,1\}^M}{\arg\max} \log p(\mathbf{s}|\mathbf{y}), \tag{2}$$

where $p(\mathbf{s}|\mathbf{y}) = \int_{\mathbf{x}} p(\mathbf{x}, \mathbf{s}|\mathbf{y}) d\mathbf{x}$. To tackle problem (2), a greedy algorithm could be used [1] to approach the solution with a succession of local decisions. In this paper, we alternatively propose a mean-field (MF) approximation of $p(\mathbf{x}, \mathbf{s}|\mathbf{y})$ which approximates $p(\mathbf{x}, \mathbf{s}|\mathbf{y})$ with a probability distribution, say $q(\mathbf{x}, \mathbf{s})$, constrained to have a "suitable" factorization while minimizing the Kullback-Leibler distance with $p(\mathbf{x}, \mathbf{s}|\mathbf{y})$. Here, $q(\mathbf{x}, \mathbf{s})$ is constrained to the structure:

$$q(\mathbf{x}, \mathbf{s}) = \prod_i q(x_i, s_i) = \prod_i q(x_i|s_i)\; q(s_i). \tag{3}$$

Then the minimization of the Kullback-Leibler distance subject to (3) can be performed by the "variational Bayes EM algorithm" (VB-EM) [3], which evaluates the $q(x_i, s_i)$'s by computing at each iteration[2]:

$$q(x_i|s_i) = \mathcal{N}(m(s_i), \Gamma(s_i)),$$

$$q(s_i) \propto \sqrt{\Gamma(s_i)}\, \exp\Big(\tfrac{1}{2}\tfrac{m(s_i)^2}{\Gamma(s_i)}\Big) \exp\Big(2s_i(b_i + \textstyle\sum_{j\neq i} w_{ij}(q(s_j=1)-1))\Big)$$

$$\text{where} \quad \Gamma(s_i) = \frac{\sigma_{x_i}^2\sigma_n^2}{\sigma_n^2 + \sigma_{x_i}^2 s_i}, \quad m(s_i) = s_i \frac{\sigma_{x_i}^2}{\sigma_n^2 + \sigma_{x_i}^2 s_i} \langle \mathbf{r}_i \rangle^T \mathbf{d}_i,$$
$$\langle \mathbf{r}_i \rangle = \mathbf{y} - \textstyle\sum_{j\neq i} q(s_j = 1)\; m(s_j = 1)\, \mathbf{d}_j.$$

[1]This distribution is equal to the one used in [1], [2] with $\mathbf{s} \in \{-1, 1\}^M$.
[2]For a sake of clarity, we drop here the iteration indices.

Compared to [1], the proposed algorithm does not make any hard decision on the SR support at each iteration but rather updates posterior probabilities. In that way, it can be seen as a *soft* process. Both algorithms have similar complexities, of order M^2 per iteration.

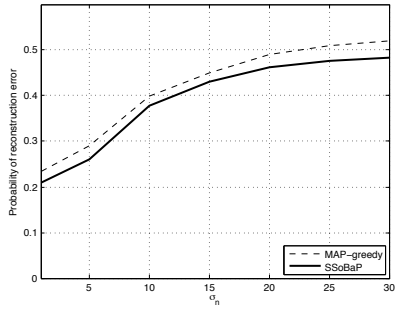
Coming back to (2), $p(\mathbf{s}|\mathbf{y})$ is simplified as $p(\mathbf{s}|\mathbf{y}) \simeq \int_{\mathbf{x}} \prod_i q(x_i, s_i) d\mathbf{x} = \prod_i q(s_i)$. We finally obtain $\forall i\ \hat{s}_i = \arg\max_{s_i\in\{0,1\}} \log q(s_i)$, which is solved by simple thresholding.

III. Experimental results

To assess the performance of the proposed algorithm, we follow the same methodology as in [1]. We generate a large number K of observations according to the model and estimate the ability of the algorithm to reconstruct the SR support via the probability

$$1 - \frac{1}{K}\sum_{k=1}^{K} \frac{\|\mathbf{s}^{(k)} \cap \hat{\mathbf{s}}^{(k)}\|_0}{\max(\|\mathbf{s}^{(k)}\|_0, \|\hat{\mathbf{s}}^{(k)}\|_0)}. \tag{4}$$

The data is generated with $N = 64$, $M = 256$, and a DCT dictionary. The Boltzmann parameters are drawn independently: the elements of $\mathbf{b}$ from $\mathcal{N}(-2.5, 1)$ and the elements of $\mathbf{W}$ from $\mathcal{U}[-0.1, 0.1]$. The standard deviations σ_{x_i} are *i.i.d.* realizations of $\mathcal{U}[15, 60]$. For each point of simulation, we run 500 trials. We adjust the final threshold at 0.25. The figure above compares 2 algorithms: "MAP-greedy", proposed in [1] and "SSoBaP" (for Structured Soft Bayesian Pursuit algorithm), proposed here. For the performance criterion considered, we can see that "SSoBaP" outperforms "MAP-greedy" over a wide range of noise variances.



IV. Conclusion

In this paper, we have shown that a MF approximation together with a VB-EM algorithm is a promising and competitive approach for the estimation of structures between atoms. To the extent of the considered criterion, the resulting algorithm is shown to outperform the baseline algorithm [1]. Complementary results, involving other performance criteria and other state-of-the-art algorithms, will be added in the final paper to confirm the relevance of this approach.

References

[1] T. Faktor, Y. C. Eldar, and M. Elad, "Exploiting statistical dependencies in sparse representations for signal recovery," Submitted to IEEE Trans. On Signal Processing.

[2] V. Cevher, M. F. Duarte, C. Hegde, and R. G. Baraniuk, "Sparse signal recovery using markov random fields," in *NIPS*, 2008.

[3] M. J. Beal and Z. Ghahramani, "The variational bayesian em algorithm for incomplete data: with application to scoring graphical model structures," *Bayesian Statistics*, vol. 7, pp. 453–463, 2003.