Robust Incorporation of Signal Predictions into the Sparse Bayesian Learning Framework

Matthew R. O'Shaughnessy, Mark A. Davenport, and Christopher J. Rozell
School of Electrical and Computer Engineering
Georgia Institute of Technology
Email: {moshaughnessy6,mdav,crozell}@gatech.edu

I. INTRODUCTION

Many signal processing problems require the estimation of sparse and time-varying signals from undersampled and noisy measurements. In the context of static sparse signals, incorporating a signal model that exploits the sparse structure present in the signal has been shown to dramatically improve estimation performance. In dynamic settings, we would like to be able to similarly exploit a priori knowledge that the signal evolves either slowly over time or based on an imperfectly known dynamics model. Most methods for incorporating both sparsity and dynamics information do so by extending variants of ℓ_1-based methods. In this work, however, we consider a sparse dynamic filtering algorithm that makes use of the *sparse Bayesian learning* (SBL) procedure, which has been shown to produce more accurate estimates than similar ℓ_1 methods [2], particularly when the dictionary contains challenging structure such as coherence and diverse column scaling [3], [4].

II. THE SBL-DF ALGORITHM

The SBL algorithm defines a hierarchical probabilistic model consisting of a Gaussian likelihood on the observations, $\boldsymbol{y} \sim \mathcal{N}(\boldsymbol{\Phi x}, \sigma^2 \boldsymbol{I})$; a zero-mean Gaussian prior on each element of x_i, $x_i \sim \mathcal{N}(0, \gamma_i)$; and a (conjugate) inverse Gamma hyperprior on each variance, $\gamma_i \sim \mathcal{IG}(a_i, b_i)$. The SBL inference procedure consists of estimating the variance parameters by maximizing $p\,(\boldsymbol{\gamma}|\boldsymbol{y})$ and then using these variances to compute the posterior estimate $\widehat{\boldsymbol{x}} = \arg\max_{\boldsymbol{x}} p\,(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{\gamma})$ (see [5], [6] for details).

The traditional SBL algorithm fixes the parameters $\{a_i, b_i\}$ of the inverse Gamma hyperpriors to make the hyperpriors either uninformative or scale-invariant. The fundamental insight of our approach is that an a priori signal estimate $\widetilde{\boldsymbol{x}}$ can be incorporated into the SBL estimator by replacing these uninformative hyperparameters with *informative* hyperpriors set using $\widetilde{\boldsymbol{x}}$. This method for incorporating prior information has been shown to be particularly robust in [7], which applied a similar strategy to the reweighted ℓ_1 procedure [8]. Intuitively, the "effective" prior formed by marginalizing over γ_i, $p\,(x_i|a_i, b_i)$, becomes wider when $|\widetilde{x}_i|$ is large, encouraging (but not forcing) x_i to be nonzero. From an optimization perspective, we can view the informative-hyperprior SBL algorithm as minimizing the objective $\ell(\boldsymbol{\gamma}) = \ell_{\text{uninf}}(\boldsymbol{\gamma}) + \ell_{\text{dyn}}(\boldsymbol{\gamma})$, where ℓ_{uninf} represents the traditional (uninformative hyperprior) SBL objective and $\ell_{\text{dyn}}(\boldsymbol{\gamma}) = 2\sum_i \left(b_i \gamma_i^{-1} - a_i \log \gamma_i^{-1}\right)$ represents the portion of the objective contributed by the informative hyperprior that we will use to incorporate dynamics information.

Our specific method for mapping the estimate $\widetilde{\boldsymbol{x}}$ to the hyperparameters a_i and b_i is performed in two steps. First, we calculate the variances $\widetilde{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} \mathbb{E}\,\|\widehat{\boldsymbol{x}} - \widetilde{\boldsymbol{x}}\|_2^2$. Intuitively, $\widetilde{\boldsymbol{\gamma}}$ represents the variances that would be used to make the estimate $\widehat{\boldsymbol{x}}$ as close as possible to the prediction $\widetilde{\boldsymbol{x}}$. This expression admits the closed-form expression $\widetilde{\gamma}_i = \widetilde{x}_i^2$ when $\boldsymbol{\Phi}^T\boldsymbol{\Phi}$ is diagonal; although this is rarely true in practice, we have found empirically that this approximation works well. Second, we choose the hyperparameters a_i and b_i so that $\widetilde{\boldsymbol{\gamma}}$ minimizes ℓ_{dyn}, which requires $\gamma_i = b_i/a_i$. Combining these expressions for $\boldsymbol{\gamma}$ and incorporating a parameter ξ to control the weight of the informative hyperpriors yields the mapping $a_i = \xi$ and $b_i = \xi\, \widetilde{x}_i^2$. When this method is used in a causal tracking context, we call the resulting algorithm SBL with dynamic filtering (SBL-DF).

Our proposed method for mapping an a priori signal estimate to the SBL probability model is flexible, allowing its use in diverse tracking scenarios. The signal estimate $\widetilde{\boldsymbol{x}}$ can be found with a dynamics model that generates either a point estimate of, or distribution over, $\boldsymbol{x}^{(t)}$. Further, the estimate can be based on either only previous time steps (producing an online algorithm) or both previous and future time steps (producing a Kalman smoothing-type framework). Although other similar methods (e.g., [9], [10]) improve performance by taking advantage of global information, the flexibility of our proposed algorithm makes it suitable for applications where this information is unavailable.

III. NUMERICAL SIMULATIONS AND DISCUSSION

The key benefit of our proposed algorithm is that, by incorporating a signal prediction into the higher-order parameters of a hierarchical probability model, the resulting estimator allows prior knowledge to be effectively exploited without significantly compromising performance when the signal estimate is inaccurate. This robustness is demonstrated in Figure 1, which shows that SBL-DF can accurately estimate $\boldsymbol{x}$ with many fewer measurements than traditional SBL requires, even when the estimate $\widetilde{\boldsymbol{x}}$ is noisy. Performance decreases as $\widetilde{\boldsymbol{x}}$ becomes less accurate, but still improves on traditional SBL, demonstrating the robustness of this approach. Here, we simulate imperfect predictions using support errors, but we have observed similar behavior with other error models such as Gaussian noise. Note that setting the tuning parameter ξ based on the accuracy of the signal prediction is necessary for optimal performance.

SBL is commonly used in settings where the dictionary contains undesirable structure such as the combination of nonuniform column scaling and coherence. Figure 2 demonstrates that SBL-DF is able to accurately reconstruct signals with larger amounts of dictionary structure than a state of the art ℓ_1-based dynamic filtering algorithm.

In addition to the improvement in estimation accuracy, incorporating a signal prediction in this way can significantly speed convergence. Figure 3 shows a decrease of approximately one order of magnitude in total computation time when using the EM iterations of [5], [6]. An additional advantage of our specific method for incorporating prior knowledge using informative hyperpriors is that it admits a simple extension of the fast marginal likelihood (FML) inference procedure of [11]. We present the details of this extension to informative hyperpriors in [1]. Figure 3 shows that this extension converges faster than the EM updates when the signal is very sparse.

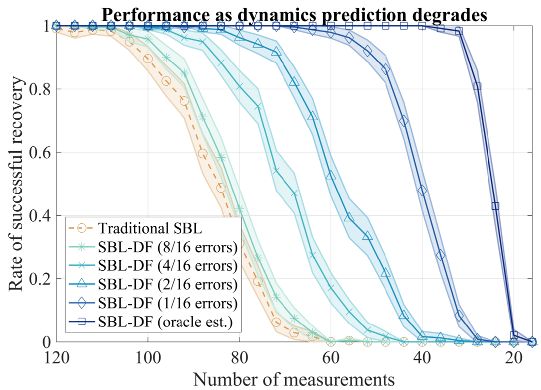

Fig. 1. Rate of successful recovery as the number of measurements M is varied. The signal $\boldsymbol{x} \in \mathbb{R}^N$, which contains $s = 16$ nonzero entries, is observed with noisy measurements $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{e}$, where dictionary $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$ is constructed with i.i.d. Gaussian entries and $\boldsymbol{e} \sim \mathcal{N}(0, 10^{-3}\boldsymbol{I})$. The signal is recovered with traditional (uninformative) SBL and with SBL-DF. SBL-DF uses a prediction $\widetilde{\boldsymbol{x}}$ generated from a corrupted version of $\boldsymbol{x}$ by randomly swapping the value of k of the 16 nonzero elements with zero-valued elements. Success is claimed when relative MSE $\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2^2 \,/\, \|\boldsymbol{x}\|_2^2 < 10^{-2}$. The tuning parameter ξ is selected independently for each error level and 240 trials are performed at each point. Error bars represent a 95% confidence interval on the rate.

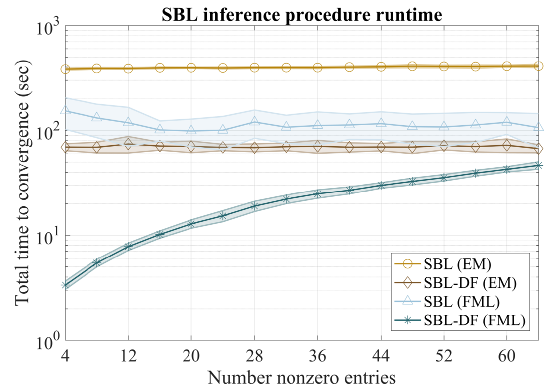

Fig. 3. Total computation time for convergence for traditional (uninformative hyperprior) SBL and SBL-DF using both the EM iterations of [5] and the FML method of [11] (which we extend to informative hyperpriors in [1]). The number of nonzero entries is varied from 4 to 64 with the problem size $N = 4096$ and $M = 1024$ fixed. Each point displays the mean and standard deviation of 24 trials. In addition to being much more efficient when the number of nonzero values is small, the FML method can also scales better with the problem size M and N; see [1] for details.

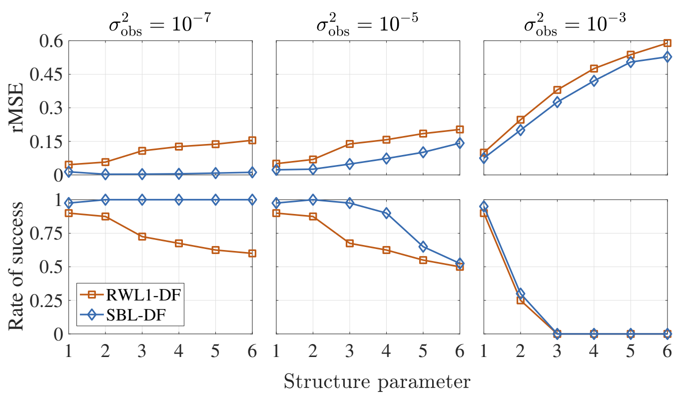

Fig. 2. Reconstruction performance in terms of relative MSE, $\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2^2 \,/\, \|\boldsymbol{x}\|_2^2 < 10^{-2}$, and rate of successful recovery as the amount of dictionary structure is varied. The "structure parameter" represents a joint measure of the column scaling and coherence between groups of columns in the dictionary (see [1] for details and a discussion of the individual effects of column scaling and coherence). The signal $\boldsymbol{x} \in \mathbb{R}^N$, which contains $s = 25$ nonzero entries, is observed with noisy measurements $\boldsymbol{y} = \boldsymbol{\Phi x} + \boldsymbol{e}$, where $\boldsymbol{\Phi} \in \mathbb{R}^{42 \times 100}$, $\boldsymbol{e} \sim \mathcal{N}(0, \sigma^2_{\text{obs}})$. The signal prediction $\widetilde{\boldsymbol{x}}$ is generated from $\boldsymbol{x}$ by swapping nonzero entries with zero-valued entries with probability $p = 0.1$, then adding white Gaussian noise with variance $\sigma^2_{\text{dyn}} = 10^{-4}$. Success is claimed when relative MSE $\|\boldsymbol{x} - \widehat{\boldsymbol{x}}\|_2^2 \,/\, \|\boldsymbol{x}\|_2^2 < 10^{-2}$. The tuning parameters (ξ for SBL-DF, and $\{\xi, \nu, \lambda_0\}$ for RWL1-DF) are selected independently for each structure and noise level, and 40 trials are performed at each point. RWL1-DF is the reweighted ℓ_1 with dynamic filtering algorithm of [7].

Acknowledgment

This work was supported in part by NSF grants CCF-1350616 and CCF-1409422, James S. McDonnell Foundation grant number 220020399, a gift from the Alfred P. Sloan Foundation, and the National Defense Science and Engineering Graduate Fellowship.

References

[1] M. R. O'Shaughnessy, M. A. Davenport, and C. J. Rozell, "Sparse Bayesian learning with dynamic filtering for inference of time-varying sparse signals," *arXiv:1902.05362 [eess.SP]*, Feb. 2019.

[2] D. Wipf and S. Nagarajan, "Iterative reweighted l1 and l2 methods for finding sparse solutions," *IEEE J. Sel. Top. Signal Process.*, vol. 4, no. 2, pp. 317–329, Apr. 2010.

[3] D. P. Wipf, "Sparse estimation with structured dictionaries," in *Proc. Adv. in Neural Inf. Process. Syst. (NeurIPS)*, Granada, Spain, Dec. 2011.

[4] Z. Zhang, "Comparison of Sparse Signal Recovery Algorithms with Highly Coherent Dictionary Matrices: The Advantage of T-MSBL," 2012.

[5] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, no. Jun, pp. 211–244, 2001.

[6] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[7] A. S. Charles, A. Balavoine, and C. J. Rozell, "Dynamic filtering of time-varying sparse signals via l1 minimization," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5644–5656, Nov. 2016.

[8] E. J. Candès, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l1 minimization," *J. Fourier Anal. Appl.*, vol. 14, no. 5-6, pp. 877–905, Dec. 2008.

[9] Z. Zhang and B. D. Rao, "Iterative reweighted algorithms for sparse signal recovery with temporally correlated source vectors," in *2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2011, pp. 3932–3935.

[10] ——, "Exploiting correlation in sparse signal recovery problems: Multiple measurement vectors, block sparsity, and time-varying sparsity," in *Proc. ICML Work. on Structured Sparsity: Learning and Inference*, Bellevue, WA, USA, May 2011.

[11] M. E. Tipping and A. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proc. Int. Work. on Artificial Intell. and Stat. (AISTATS)*, Key West, FL, USA, Jan. 2003.