

TRACKING DYNAMIC SPARSE SIGNALS USING HIERARCHICAL BAYESIAN KALMAN FILTERS

Evripidis Karseras, Kin Leung, and Wei Dai

Department of Electrical and Electronic Engineering
Imperial College, London, UK
{e.karseras11, kin.leung, wei.dai1}@imperial.ac.uk

ABSTRACT

In this work we are interested in the problem of reconstructing time-varying signals for which the support is assumed to be sparse. For a single time instance it is possible to reconstruct the original signal efficiently by employing a suitable algorithm for sparse signal recovery, given the sparsity level of the signal. In the case of time-varying *sparse* signals the sparsity level is not necessarily known *a-priori*. Furthermore conventional tracking by *Kalman filtering* fails to promote sparsity. Instead, a *hierarchical Bayesian model* is used in the tracking process which succeeds in modelling sparsity. One theorem is provided that extends previous work by providing some more general results. A second theorem gives the conditions under which all sparse signals are recovered exactly. It is demonstrated that the proposed method succeeds in recovering time-varying sparse signals with greater accuracy than the classic Kalman filter approach.

Index Terms— Hierarchical Bayesian network, Kalman filter, time-varying sparse signals

1. INTRODUCTION

In this work we consider the problem of reconstructing time varying signals that are assumed to be sparse. In the case where temporal measurements are taken then the problem of tracking dynamic sparse signals cannot be solved with classic approaches since the underlying models do not promote sparse signals. In the last decade the field of sparse signal representation and reconstruction has gained a lot of interest in many fields mostly because many signals of interest can be sparsely represented in some basis. In many cases not only do we have to deal with sparse signals but a time-series of sparse signals which we seek to track. Examples can be found in many applications such as computer vision where tracking a set of features needs to take place, data fusion from vast wireless sensor networks and many more.

The authors would like to acknowledge the European Commission for funding SmartEN ITN (Grant No. 238726) under the Marie Curie ITN FP7 programme. This work is partially supported by the University Defence Research Centre (UDRC), UK.

For a single time instance, the problem of sparse signal reconstruction has been addressed extensively with research in the field of compressed sensing. In [1] a specific type of a Bayesian network was introduced which elegantly promotes sparsity. This learning framework namely the Relevance Vector Machine (RVM) results in highly sparse models for the input and has gained popularity in the signal processing community for its use in compressed sensing applications [2] and basis selection [3]. Sparsity is rendered possible with a hierarchy of prior distributions which intuitively confines the space of all possible states. The key fact behind this technique is that it provides estimates on full distributions. Non-Bayesian sparse recovery algorithms do not take into account the signal's statistics making their use in tracking sparse signals difficult. The resulting statistical information can be used to make predictions for future states without the risk of them not being sparse. In addition there is no need for a predetermined level of sparsity since the number of active components is automatically inferred. This makes the method even more appealing for use in an on-line tracking algorithm. Subsequent work in [4, 5] has shown that the properties of this model also allow for efficient implementation.

For dynamic signals the system's state can be tracked by the Kalman filter with great accuracy. Unfortunately this approach is not fit for sparse signals. The Kalman filter requires that the mean value of the predicted states to be equal with the mean value of the actual state while minimising variation. Bearing this in mind the original principle must be extended. By examining the formulation of this estimator it becomes evident that the model must be revisited. Previous work in [6, 7] employs a modified Kalman filter which does not take into account a sparse model and also requires to specify a number of sensitive external parameters.

The Bayesian network of the RVM is employed to extend the data model adopted in traditional tracking so as to admit sparse solutions. The automatic determination of the active components solves the problem of having to assume fixed sparsity levels. This is of great importance since in a real-time tracking scenario the sparsity level constantly changes. The resulting statistical information is then incorporated into

updating predictions and thus making sparse state estimates. One theorem is presented which extends previous work in [3] by providing some more general results on the associated cost function. A second theorem shows that under certain conditions all sparse signals can be recovered exactly. Simulations show that the proposed method achieves lower reconstruction error than the classic Kalman filter approach.

In the next section we briefly mention the basic properties of the Kalman filter. Also the key properties of sparse Bayesian learning introduced in [1] are outlined. A discussion is made regarding previous work and the main differences with ours. In Section 3 the connection to the initial problem is highlighted and the revised prediction and update steps are given along with the two theorems. Finally simulation results on a toy problem in Section 4 highlight the performance of this amalgamation.

2. PRELIMINARIES

The system model on which this work is based is described by the following two equations:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \mathbf{q}_t, \quad (1)$$

$$\mathbf{y}_t = \mathbf{\Phi}_t \mathbf{x}_t + \mathbf{n}_t. \quad (2)$$

Vectors $\mathbf{x}_t, \mathbf{y}_t$ denote the system's state and measurements respectively. The state's innovation is modelled by process \mathbf{q}_t and measurement noise by \mathbf{n}_t . In this work and without loss of generality we also assume that $\mathbf{x}_t \in \mathbb{R}^n$ is *s-sparse* and design matrix $\mathbf{\Phi}_t \in \mathbb{R}^{m \times n}$.

In this section the main aspects of Kalman filtering and sparse Bayesian learning (SBL) are briefly presented. The classic state-space model in Equations (1) and (2) assumes Gaussian distributions on the model parameters and thus is inappropriate for sparse signals. On the other hand SBL achieves to promote sparsity for a single time-instance of Equation (2). In Section 3 the connection between the two is presented.

2.1. The Kalman filter and the related approaches

The most popular technique used for tracking a dynamic signal is the Kalman filter. It is based on the Gaussian assumption, that is: $p(\mathbf{x}_t | \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t, \mathbf{Q}_t)$ and $p(\mathbf{y}_t | \mathbf{x}_t) = \mathcal{N}(\mathbf{\Phi}_t \mathbf{x}_t, \sigma^2 \mathbf{I})$ while $p(\mathbf{q}_t) = \mathcal{N}(\mathbf{0}, \mathbf{Q}_t)$ and $p(\mathbf{n}_t) = \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. To track the statistics of a dynamic signal it is sufficient to iterate a two step procedure: the Kalman filter prediction and update steps. The prediction step calculates the parameters of $p(\mathbf{x}_t | \mathbf{y}_{t-1})$ while the update step evaluates those of $p(\mathbf{x}_t | \mathbf{y}_t)$. It can be shown that both distribution functions are Gaussian. As a result, not only one can obtain the globally optimal estimate of \mathbf{x}_t in terms of mean squared error (MSE), but also track the full distribution exactly. However, and as can be easily verified, the optimal estimate of the

standard Kalman filter is typically not sparse. Model modifications must be done for tracking dynamic sparse signals.

There are several approaches in which the Kalman filter framework is adopted but slightly modified to admit sparse solutions. The essential idea behind [6] and [7] is to apply thresholds that enforce sparsity. Work in [8] adopts a probabilistic model but signal amplitudes and support are estimated separately. Techniques presented in [9] use prior sparsity knowledge into the tracking process. All these approaches typically require a number of parameters to be pre-set. It also remains unclear how these methods perform towards model and parameter mismatch.

2.2. RVM for sparse signal reconstruction

For a more elaborate analysis the interested reader can refer to [1, 5]. The following discussion refers to Equation (2) for just one time instance. Note that time index t is temporarily dropped.

SBL was proposed in [2] to solve the problem of sparse signal reconstruction. It was underscored in subsection 2.1 that if $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q})$ where \mathbf{Q} is full rank, then the minimum MSE solution to Equation (2) provided by the Kalman optimal estimator is typically not sparse. In SBL this issue is addressed by introducing individual hyper-parameters $\boldsymbol{\alpha}$ to control the variance of each component x_i :

$$p(\mathbf{x} | \boldsymbol{\alpha}) = \prod_{i=1}^n \mathcal{N}(0, \alpha_i^{-1}) = \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1})$$

where matrix $\mathbf{A} = \text{diag}([\alpha_1, \dots, \alpha_n])$. Driving $\alpha_i = +\infty$ results in $p(x_i | \alpha_i) = \mathcal{N}(0, 0)$ which means it is *a-posteriori* certain that $x_i = 0$. Hence, the reconstruction problem is then changed to finding the maximum likelihood solution of $\boldsymbol{\alpha}$ for the given measurements \mathbf{y} . The explicit form of the likelihood $p(\mathbf{y} | \boldsymbol{\alpha}, \sigma^2)$ was derived in [1]. A set of fast algorithms to estimate $\boldsymbol{\alpha}$ and hence \mathbf{x} are proposed in [5]. Even though the algorithms in [5] are not guaranteed to produce globally optimal solutions, they perform extremely well for many practical scenarios.

At this point we recognise the elegance of this method. There are no control parameters to be manually set, like a pre-determined level of sparsity. A stopping criterion is needed for the optimisation process and this can be set to some safe threshold so as to not affect the validity of the process. This is of great importance for our method since the sparsity level of \mathbf{x}_t is unknown and varying. Furthermore, the unknown state vector \mathbf{x} is Gaussian distributed (conditioned on a given $\boldsymbol{\alpha}$). This allows us to incorporate the SBL framework into the Kalman filter.

3. HIERARCHICAL BAYESIAN KALMAN FILTER

For the purpose of tracking dynamic sparse signals, the principles behind the Kalman filter and SBL are combined. In

particular, Equations (1) and (2) are still used to model the dynamic system. The measurement noise is assumed to be Gaussian with known covariance, i.e. $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$. Differently from the standard Kalman filter where the state innovation \mathbf{q}_t is Gaussian with a given covariance matrix \mathbf{Q} , we assume that $\mathbf{q}_t \sim \mathcal{N}(0, \mathbf{A}_t^{-1})$ where $\mathbf{A}_t = \text{diag}(\alpha_t) = \text{diag}([\alpha_1, \dots, \alpha_n]_t)$ and hyper-parameters α_i are not priorly known and will have to be learned from the observation vector \mathbf{y}_t .

Similar to the standard Kalman filter the two steps of prediction and update still need to be performed at each time instance. In the prediction step, one has:

$$\begin{aligned} \boldsymbol{\mu}_{t|t-1} &= \boldsymbol{\mu}_{t-1}, \quad \boldsymbol{\Sigma}_{t|t-1} = \boldsymbol{\Sigma}_{t-1} + \mathbf{A}_t^{-1}, \\ \mathbf{y}_{t|t-1} &= \boldsymbol{\Phi}_t \boldsymbol{\mu}_{t|t-1}, \quad \mathbf{y}_{e,t} = \mathbf{y}_t - \mathbf{y}_{t|t-1}. \end{aligned} \quad (3)$$

where the notation $t|t-1$ means prediction at time instance t for measurements up to time instance $t-1$. In the update step, one computes:

$$\begin{aligned} \boldsymbol{\mu}_{t|t} &= \boldsymbol{\mu}_{t|t-1} + \mathbf{K}_t \mathbf{y}_{e,t}, \quad \boldsymbol{\Sigma}_{t|t} = (\mathbf{I} - \mathbf{K}_t \boldsymbol{\Phi}_t) \boldsymbol{\Sigma}_{t|t-1} \\ \mathbf{K}_t &= \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{\Phi}_t^T (\sigma^2 \mathbf{I} + \boldsymbol{\Phi}_t \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{\Phi}_t^T)^{-1} \end{aligned}$$

In addition to these steps is the additional step of learning the hyper-parameters α_t . From Equation (3), $\mathbf{y}_{e,t} = \boldsymbol{\Phi}_t \mathbf{q}_t + \mathbf{n}_t$ where a sparse \mathbf{q}_t is preferred to produce a sparse \mathbf{x}_t . As per the analysis in [1, 5], maximising the likelihood $p(\mathbf{y}_t | \alpha_t)$ is equivalent to minimising the following cost function:

$$\mathcal{L}(\alpha_t) = \log |\boldsymbol{\Sigma}_\alpha| + \mathbf{y}_{e,t}^T \boldsymbol{\Sigma}_\alpha^{-1} \mathbf{y}_{e,t}, \quad (4)$$

where $\boldsymbol{\Sigma}_\alpha = \sigma^2 \mathbf{I} + \boldsymbol{\Phi}_t \mathbf{A}_t^{-1} \boldsymbol{\Phi}_t^T$. The algorithms described in [5] can be applied to estimate α_t . Note that the cost function $\mathcal{L}(\alpha)$ is not convex. The obtained estimate α_t is generally sub-optimal (details on the estimation of the globally optimal α_t are given in Section 3.1.) Nevertheless, this sub-optimal solution is proved to be very useful in practice.

The proposed HB-Kalman has several advantages. Firstly, by adopting the system model described in Equations (1) and (2), one can track the mean and covariance of the state vector \mathbf{x}_t . Secondly, the employment of hyper-parameters to model state innovation promotes sparsity. Thirdly, different from previous work in [6, 7, 8, 9], the proposed model only involves one parameter σ^2 that needs to be manually adjusted. In principle, this parameter can also be learned from the measurements \mathbf{y}_t [1]. In our model, we let σ^2 be priorly known since this is the case with most real-world applications (most commercial sensors have to label their noise floor). Finally, as we shall show in Section 3.1, the proposed model allows certain performance guarantees (global optimality under certain conditions).

3.1. Performance Guarantees

This subsection discusses the performance guarantees of RVM at a given time instant. For convenience, the subscript t

is dropped and we focus on the model specified in Equation (2) where $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{A}^{-1})$. In the literature the RVM was analysed in [3]. It had been proven that a maximally sparse solution of $\mathbf{y} = \boldsymbol{\Phi} \mathbf{x}$ attains the global minimum of the cost function. However, the analysis did not specify the conditions to avoid local minima. By contrast, we provide more refined analysis and derive the conditions under which the original RVM algorithm [5] converges to the global minimum. Due to space constraints, only the main results are presented and all detailed proofs are delayed to the journal version of this paper.

In the performance analysis, we follow [3] by driving noise variance $\sigma^2 \rightarrow 0$, which corresponds to the noiseless setting. The following Theorem specifies the behaviour of the cost function $\mathcal{L}(\alpha)$.

Theorem 1. *For any given α , define the set $\mathcal{I} \triangleq \{1 \leq i \leq n : 0 < \alpha_i < \infty\}$. Then it holds that*

$$\lim_{\sigma^2 \rightarrow 0} \sigma^2 \mathcal{L}(\alpha) = \left\| \mathbf{y} - \boldsymbol{\Phi}_{\mathcal{I}} \boldsymbol{\Phi}_{\mathcal{I}}^\dagger \mathbf{y} \right\|_2^2, \quad (5)$$

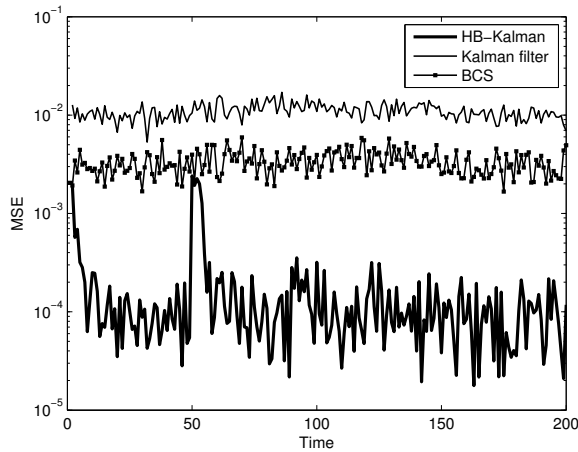
where $\boldsymbol{\Phi}_{\mathcal{I}}$ is a sub-matrix of $\boldsymbol{\Phi}$ formed by the columns indexed by \mathcal{I} , and $\boldsymbol{\Phi}_{\mathcal{I}}^\dagger$ denotes the pseudo-inverse of $\boldsymbol{\Phi}$. Furthermore, if $|\mathcal{I}| < m$ and $\mathbf{y} \in \text{span}(\boldsymbol{\Phi}_{\mathcal{I}})$, then $\mathcal{L}(\alpha) \rightarrow -\infty$ and $\sigma^2 \mathcal{L}(\alpha) \rightarrow 0$ as $\sigma^2 \rightarrow 0$.

Two observations can be obtained. Firstly, the scenarios analysed in [3] are special cases of Theorem 1 where $\mathcal{L}(\alpha) \rightarrow -\infty$. Secondly, a proper scaling of the cost function gives the squared ℓ_2 -norm of the reconstruction error. Reconstruction is then equivalent to recovering a support set that minimises the reconstruction distortion. Note that this principle is exactly the one behind many greedy algorithms including OMP [10] and subspace pursuit [11]. Theorem 1 suggests certain connections between SBL and greedy algorithms.

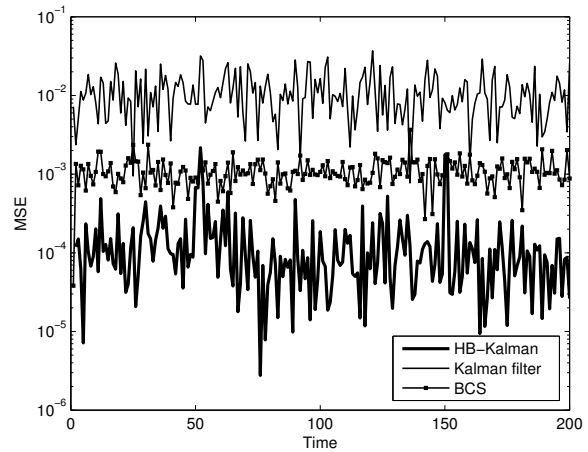
The connection becomes more clear when studying the RVM inference algorithms in [5]. Let us consider the case that $\sigma^2 = 0$. It can be shown (details are given in the journal version of this paper) that one of the sequential algorithms in [5] closely resembles the OMP technique for standard CS reconstruction [10]. As a consequence, a global performance guarantee, similar to the one in [10], can be obtained (see Theorem 2 below). To the authors' knowledge, this is the first result to prove the global performance guarantee for the RVM inference algorithm.

Theorem 2. *Assume the noiseless setting $\mathbf{y} = \boldsymbol{\Phi} \mathbf{x}$ where the columns of $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}$ are normalised, i.e. $\phi_i^T \phi_i = 1$ for all $1 \leq i \leq n$. Suppose that the matrix $\boldsymbol{\Phi}$ satisfies the incoherence condition that $|\phi_i^T \phi_j| < 0.375/s$ for all $1 \leq i \neq j \leq n$. Then the sequential algorithm to solve the RVM inference problem [5] reconstructs all s -sparse signals exactly.*

The form of the cost function in (5) also suggests that it is possible to design inference techniques based on advanced CS reconstruction algorithms. The detailed algorithm design



(a) Number of measurements $m = 100$



(b) Number of measurements $m = 6$.

Fig. 1: Tracking performance comparison.

and the corresponding performance guarantees are discussed in the journal version of this paper.

4. EMPIRICAL RESULTS

In this section a comparison is made between the original Kalman filter and the proposed method namely Hierarchical Bayesian Kalman filter (HB-Kalman). The stand-alone implementation of Bayesian compressed sensing (BCS) as presented in [2] is also tested in order to assess the quality of reconstruction when conditioning on past estimates.

A toy problem is constructed in order to obtain reconstruction results. Sparse signal $\mathbf{x}_t \in \mathbb{R}^n$ is chosen to be sparse in its natural basis. The indices i of the non-zero entries are chosen uniformly from $[1, n]$ where the n is the dimensionality of \mathbf{x}_t and is chosen to be $n = 512$. It is also assumed that $q_i \sim \mathcal{N}(0, \sigma_q^2) \forall x_i \neq 0$ with $\sigma_q^2 = 0.1$.

The simulation time for this experiment is $T = 200$ time instances. At two arbitrarily chosen time instances $t = 50$ and $t = 150$ a change in the support of \mathbf{x}_t is introduced. A non-zero component is added to the support of \mathbf{x}_{50} and a non-zero component is removed from the support of \mathbf{x}_{150} . The indices of the elements to be added and removed are also chosen at random. Apart from these two time instances the support of \mathbf{x}_t remains stationary. At $t = 1$ the support is initialised with 5 non-zero components. Noise variance is set to $\sigma^2 = 0.01$ for the entire simulation time.

Two scenarios are considered. In the first scenario noisy measurements \mathbf{y}_t are then taken by choosing the design matrix $\Phi_t \in \mathbb{R}^{100 \times 512}$ as described in Section 2.1 which is resampled at each time instance. The number of measurements at each time instance remains constant. In Figure 1(a) the mean squared reconstruction error is plotted against time for

each of the three reconstruction methods for the first scenario. It is evident that the error levels are much lower for the HB-Kalman filter when compared to the conventional Kalman filter, direct consequence of the assumed sparse model. By comparing to the repeated application of the BCS method, i.e assuming independent, identically distributed data, we see that incorporating statistical information from previous estimates results in lower reconstruction error.

In the second scenario we assume that the initial state of the system x_0 is known *a-priori* while the number of measurements taken at each time instance now reduces to $m = 6 < 2 \times 5$. This represents a difficult case since the number of measurements is less than twice the number of non-zero components which is 5 and tests the robustness of the proposed method. By observing Figure 1(b) we see that HB-Kalman still performs optimally compared to the other two methods. It is shown that given statistical information from an earlier time instance the proposed method manages to retain its performance. We also notice that error peaks have become more abrupt in the case of the Kalman filter compared to the first test scenario. The HB-Kalman approach manages to recover the correct support set in all cases.

5. CONCLUSION

We proposed a method for tracking dynamic sparse signals; a problem which can not be solved with traditional approaches like the Kalman filter. The theoretical part shows that certain connections exist with well known compressed sensing algorithms and that it is possible to incorporate previous knowledge in the tracking process. The technique is tested against some synthetic scenarios and is empirically shown to achieve better performance than traditional methods.

6. REFERENCES

- [1] Michael E. Tipping, “The relevance vector machine,” 2000.
- [2] Shihao Ji, Ya Xue, and L. Carin, “Bayesian compressive sensing,” *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, june 2008.
- [3] 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.
- [4] Anita C. Faul and Michael E. Tipping, “Analysis of sparse Bayesian learning,” in *Advances in Neural Information Processing Systems 14*. 2001, pp. 383–389, MIT Press.
- [5] M.E. Tipping, A.C. Faul, et al., “Fast marginal likelihood maximisation for sparse Bayesian models,” in *International workshop on artificial intelligence and statistics*. Jan, 2003, vol. 1.
- [6] N. Vaswani, “Kalman filtered compressed sensing,” in *IEEE International Conference on Image Processing*, oct. 2008, pp. 893–896.
- [7] A. Carmi, P. Gurfil, and D. Kanevsky, “Methods for sparse signal recovery using Kalman filtering with embedded pseudo-measurement norms and quasi-norms,” *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2405–2409, april 2010.
- [8] J. Ziniel, L.C. Potter, and P. Schniter, “Tracking and smoothing of time-varying sparse signals via approximate belief propagation,” in *Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*. IEEE, 2010, pp. 808–812.
- [9] A. Charles, M.S. Asif, J. Romberg, and C. Rozell, “Sparsity penalties in dynamical system estimation,” in *Conference on Information Sciences and Systems (CISS)*. IEEE, 2011, pp. 1–6.
- [10] J.A. Tropp, “Greed is good: algorithmic results for sparse approximation,” *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, oct. 2004.
- [11] Wei Dai and Olgica Milenkovic, “Subspace pursuit for compressive sensing signal reconstruction,” *IEEE Trans. Inform. Theory*, vol. 55, pp. 2230–2249, 2009.