Tracking Dynamic Sparse Signals with Kalman Filters: A less greedy approach

***Abstract*—Many techniques based on the traditional Kalman filter perform optimally for conventional signals but they tend to fail when it comes to recovering dynamic sparse signals. In this paper a method to solve the problem of tracking dynamic sparse signals is described. The basic idea behind the proposed method is the employment of a hierarchical Bayesian network to model the system dynamics which does not fail to capture the inherent sparsity of the data in contrast to the traditional state-space model. The probabilistic model provides all the statistical information necessary to perform sparsity-aware predictions and updates in the Kalman filter steps. In the core of this work also lies a theorem which provides the motivation and theoretical tools to construct a more efficient Bayesian inference algorithm, as opposed to previously proposed greedy methods found in the bibliography. Simulations show how the proposed method outperforms the traditional Kalman filter when it comes to dynamic sparse signals and also how the revised inference algorithm, termed here Bayesian Subspace Pursuit (BSP) achieves better performance than the previously proposed greedy methods.**

I. Introduction

The Kalman filter has played an important role in the area of linear dynamic system modelling and has been the workhorse in both practical and theoretic scenarios. Unfortunately the escalating trend towards sparse signal representation has rendered this optimal estimator to be useless when it comes to tracking *dynamic sparse signals*. It is not difficult to show that the estimation process behind the Kalman filter is not fit for sparse signals. Probably the most intuitive way to explain this is by merely examining the Bayesian formulation of the filter, where the *prior* distribution placed over the system's observations is Gaussian, thus not placing any sparse constraint over the space of all possible solutions.

One can find in the bibliography cases where the Kalman filter is externally modified to admit sparse solutions. The idea in [1] and [2] is to enforce sparsity by applying certain threshold operators. Work in [3] adopts a probabilistic model but signal amplitudes and support are estimated separately. Finally, the techniques presented in [4] 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.

For a single time instance of the sparse reconstruction problem, the Relevance Vector Machine (RVM) introduced in [5] was used with great success in compressed sensing applications [6] and basis selection [7]. The hierarchical Bayesian network behind the RVM achieves highly sparse models for the data not only providing estimates for sparse signals but estimates on their full posterior distributions as well. This is of great importance since it provides all the necessary statistical information to use in the prediction step of the tracking process. Additionally, the inference procedure used in this framework allows for automatic determination of the active components hence there is no need for a pre-determined level of sparsity or threshold. This is an appealing attribute for an on-line tracking algorithm.

In this work the hierarchical Bayesian network is employed to extend the state-space model adopted in the traditional Kalman filter. This way the problem of modelling sparsity is tackled efficiently. The resulting statistical information from the inference procedure is then incorporated in the Kalman filter prediction and update steps thus producing sparsity-aware state estimates. In the core of this work we present a theorem which provides the motivation to construct more efficient inference algorithms since the original algorithm suggests a pure greedy approach. As an answer to this a less greedy algorithm is derived based on state of the art compressed sensing reconstruction algorithms. Empirical results show how the proposed methods outperform the traditional Kalman filter and the previously proposed inference mechanisms.

In Section II we present the basic idea for amalgamating the inference procedure of the RVM in the Kalman filter along with the revised prediction and update steps of the filter, termed here Hierarchical Kalman filter (HB-Kalman). Moving to Section III we give the aforementioned theorem and explain how it motivates us to improve upon previous techniques. Additionally we provide the steps for a revised inference algorithm based on the Subspace Pursuit reconstruction algorithm in [9], termed here Bayesian Subspace Pursuit (BSP). In Section IV simulations show how the proposed methods perform in some synthetic scenarios.

II. Hierarchical Bayesian Kalman Filter

The system model on which this discussion builds is described by the following equations:

$$\boldsymbol{x}_t = \boldsymbol{x}_{t-1} + \boldsymbol{q}_t, \tag{1}$$

$$\boldsymbol{y}_t = \boldsymbol{\Phi}_t \boldsymbol{x}_t + \boldsymbol{n}_t. \tag{2}$$

where vectors $\boldsymbol{x}_t, \boldsymbol{y}_t$ denote the system's state and observations respectively. The state innovation and observation noise processes are modelled by $\boldsymbol{q}_t$ and $\boldsymbol{n}_t$ respectively. Furthermore we assume that signal $\boldsymbol{x}_t \in \mathbb{R}^n$ is sparse in some transform domain and that the entries of design matrix $\boldsymbol{\Phi}_t \in \mathbb{R}^{m \times n}$ are independently identically distributed Gaussian random variables.

Based on the Gaussian assumption: $p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}) = \mathcal{N}(\boldsymbol{x}_{t-1}, \boldsymbol{Q}_t)$ and $p(\boldsymbol{y}_t|\boldsymbol{x}_t) = \mathcal{N}\left(\boldsymbol{\Phi}\boldsymbol{x}_t, \sigma^2\boldsymbol{I}\right)$ while $p(\boldsymbol{q}_t) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}_t)$ and $p(\boldsymbol{n}_t) = \mathcal{N}\left(\boldsymbol{0}, \sigma^2\boldsymbol{I}\right)$, the Kalman filter continuously alternates between the prediction and update step. The prediction step calculates the parameters of $p(\boldsymbol{x}_t|\boldsymbol{y}_{t-1})$ while the update step evaluates those of $p(\boldsymbol{x}_t|\boldsymbol{y}_t)$.

The mean squared error solutions given to Equation (2) for a single time instance in general are not sparse. To alleviate this problem, the key idea behind Sparse Bayesian Learning (SBL) [5] is employed. An additional level of parameters $\boldsymbol{\alpha}$ is introduced to control the variance of each component x_i:

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

where matrix $\boldsymbol{A} = \mathrm{diag}\left([\alpha_1, \cdots, \alpha_n]\right)$. By driving $\alpha_i = +\infty$ means that $p(x_i|\alpha_i) = \mathcal{N}(0, 0)$ hence it is certain that $x_i = 0$. What remains is to find the maximum likelihood solution of $\boldsymbol{\alpha}$ for the given observation vector $\boldsymbol{y}$. The explicit form of the likelihood function $p\left(\boldsymbol{y}|\boldsymbol{\alpha}, \sigma^2\right)$ was derived in [5] and a set of fast algorithms to estimate $\boldsymbol{\alpha}$ and consequently $\boldsymbol{x}$ are proposed in [10].

Finally the principles behind the Kalman filter and SBL are put together. In particular, Equations (1) and (2) are still used to model the dynamic system. The measurement noise is chosen to be Gaussian with known covariance, i.e. $\boldsymbol{n} \sim \mathcal{N}\left(0, \sigma^2\boldsymbol{I}\right)$. Here it is assumed that the state innovation process is given by $\boldsymbol{q}_t \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{A}_t^{-1}\right)$, with $\boldsymbol{A}_t = \mathrm{diag}\left(\boldsymbol{\alpha}_t\right) = \mathrm{diag}\left([\alpha_1, \cdots, \alpha_n]_t\right)$ and the hyper-parameters α_i are to be learned from the data, as opposed to the traditional Kalman filter where the covariance matrix $\boldsymbol{Q}$ of $\boldsymbol{q}_t$ is given.

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

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

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:

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

Differently from the standard Kalman filter, one has to perform the additional step of learning the hyper-parameters $\boldsymbol{\alpha}_t$. From Equation (3) we get $\boldsymbol{y}_{e,t} = \boldsymbol{\Phi}_t\boldsymbol{q}_t + \boldsymbol{n}_t$ where a sparse $\boldsymbol{q}_t$ is preferred to produce a sparse $\boldsymbol{x}_t$. Following the analysis in [5] and [10], maximising the likelihood $p(\boldsymbol{y}_t|\boldsymbol{\alpha}_t)$ is equivalent to minimising the following cost function:

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

where $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \sigma^2\boldsymbol{I} + \boldsymbol{\Phi}_t\boldsymbol{A}_t^{-1}\boldsymbol{\Phi}_t^T$. The algorithms described in [10] can be applied to estimate $\boldsymbol{\alpha}_t$. Note that the cost function $\mathcal{L}(\boldsymbol{\alpha})$ is not convex. The obtained estimate $\boldsymbol{\alpha}_t$ is generally sub-optimal and details on the estimation of the globally optimal $\boldsymbol{\alpha}_t$ are given in the next section.

III. Bayesian Subspace Pursuit

Here we discuss the performance guarantees of the inference procedure at a given time instant and demonstrate how it can be further improved. For convenience, subscript t is dropped and focus is turned to Equation (2) where $\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{A}^{-1}\right)$. In order to analyse and possibly improve a certain algorithm one has to account for the case where signal perturbations are absent. In the literature this was analysed in [7] for the purpose of Basis Selection. It was also proved in [7] that a maximally sparse solution of $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}$ attains the global minimum of the cost function. However, the analysis did not specify the conditions to avoid local minima. By contrast, we provide a more refined analysis. Due to space constraints, only the main results are presented.

We follow [7] by driving the noise variance $\sigma^2 \to 0$, which corresponds to the noiseless case. The following Theorem specifies the behaviour of the cost function $\mathcal{L}(\boldsymbol{\alpha})$.

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

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

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 $\boldsymbol{y} \in \mathrm{span}(\boldsymbol{\Phi}_{\mathcal{I}})$, then $\mathcal{L}(\boldsymbol{\alpha}) \to -\infty$ and $\sigma^2\mathcal{L}(\boldsymbol{\alpha}) \to 0$ as $\sigma^2 \to 0$.

Two observations can be obtained. Firstly, the scenarios analysed in [7] can be seen as special cases of Theorem 1 where $\mathcal{L}(\boldsymbol{\alpha}) \to -\infty$. 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. This principle is the same as the one behind many greedy algorithms including OMP [8] and Subspace Pursuit [9]. Theorem 1 suggests such connections.

Based on the results from Theorem 1 the key quantities concerning the algorithms described in [10] are scaled accordingly, as the Theorem suggests. The original formulae can be found in [10] while the revised ones are given below:

$$\sigma^{-2}\boldsymbol{\Sigma}_x = \left(\sigma^2\boldsymbol{A}_{\mathcal{I}} + \boldsymbol{\Phi}_{\mathcal{I}}^T\boldsymbol{\Phi}_{\mathcal{I}}\right)^{-1}, \quad \boldsymbol{\mu}_x = \sigma^{-2}\boldsymbol{\Sigma}_x\boldsymbol{\Phi}_{\mathcal{I}}^T\boldsymbol{y}$$

$$\sigma^2 s_i = \boldsymbol{\phi}_i^T\left(\sigma^2\boldsymbol{C}_{-i}^{-1}\right)\boldsymbol{\phi}_i, \quad \sigma^2 q_i = \boldsymbol{\phi}_i^T\left(\sigma^2\boldsymbol{C}_{-i}^{-1}\right)\boldsymbol{y}$$

where $\sigma^2\boldsymbol{C}_{-i}^{-1} = \boldsymbol{I} - \boldsymbol{\Phi}_{\mathcal{I}-i}\left(\sigma^2\boldsymbol{A}_{\mathcal{I}-i} + \boldsymbol{\Phi}_{\mathcal{I}-i}^T\boldsymbol{\Phi}_{\mathcal{I}-i}\right)^{-1}\boldsymbol{\Phi}_{\mathcal{I}-i}^T$. Subscript $\mathcal{I}$ denotes the subset of columns of $\boldsymbol{\Phi}$ for which $0 < \alpha_i < +\infty$. Notation $\mathcal{I} - i$ means the removal of index i from $\mathcal{I}$. Subsequently the expression for the hyper-parameters becomes:

$$\alpha_i = \frac{(\sigma^2 s_i)^2}{(\sigma^2 q_i)^2 - \sigma^4 s_i}$$

Finally the scaled cost function becomes:

$$\begin{aligned} \sigma^2\mathcal{L} =& \sigma^2\log\left|\sigma^2\boldsymbol{I} + \boldsymbol{\Phi}_{\mathcal{I}}\boldsymbol{A}_{\mathcal{I}}^{-1}\boldsymbol{\Phi}_{\mathcal{I}}^T\right| + \\ & \boldsymbol{y}^T\left(\boldsymbol{I} - \boldsymbol{\Phi}_{\mathcal{I}}(\sigma^2\boldsymbol{A}_{\mathcal{I}} + \boldsymbol{\Phi}_{\mathcal{I}}^T\boldsymbol{\Phi}_{\mathcal{I}})^{-1}\boldsymbol{\Phi}_{\mathcal{I}}^T\right)\boldsymbol{y}. \end{aligned}$$

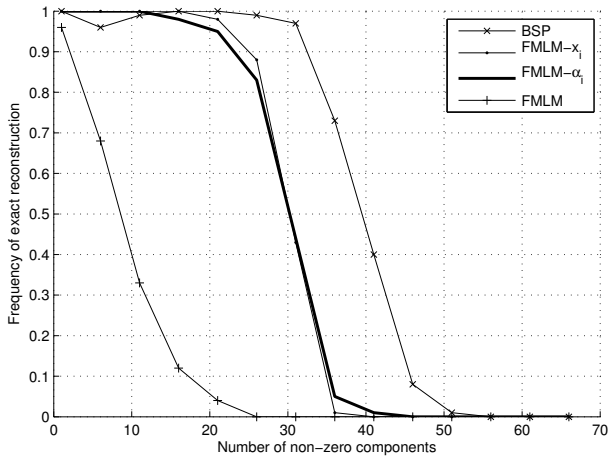

Fig. 1: Exact reconstruction rates for $m = 128, n = 256$

Based on these expressions we can re-formulate the algorithm termed Fast Marginal Likelihood Maximisation (FMLM) found in [10] to facilitate both assumptions of presence and absence of noise. Furthermore, in order to improve performance, the selection of basis vectors is based on the values of components $\boldsymbol{x}_i$ instead of the hyper-parameters α_i. This algorithm, termed henceforth FMLM-$\boldsymbol{x}_i$ is compactly presented below:

Algorithm 1 FMLM-$\boldsymbol{x}_i$

Input: $\boldsymbol{\Phi}, \boldsymbol{y}, \sigma^2$
Initialise:
- Initialise iteration counter, $k = 1$.
- $T_k = \{\text{index } i \in [1, n] \text{ for which } |\boldsymbol{\phi}_i^T \boldsymbol{y}| \text{ is maximised}\}$.

Iteration:
- Calculate values of α_i and $[\boldsymbol{\mu}_x]_i$ for $i \in [1, n] \setminus T_k$.
- Increase counter, $k = k + 1$.
- $T_k = T_{k-1} \cup \{\text{index } i \text{ corresponding to the maximum value of } [\boldsymbol{\mu}_{\boldsymbol{x}}]_i \text{ for } i \notin T_k\}$.
- Calculate α_i for $i \in T_k$.
- Remove indices $\{i \in T_k \text{: } \alpha_i < 0, \alpha_i = +\infty\}$ from T_k.
- If $\sigma^2 \mathcal{L}_k$ has reached a steady state then calculate $\sigma^{-2}\boldsymbol{\Sigma}_x$, $\boldsymbol{\mu}_x$ for $i \in T_k$ and quit. Continue otherwise.

Output:
- Estimated support set T_k and sparse signal $\hat{\boldsymbol{x}}$ with $|T_k|$ non-zero components, $\hat{\boldsymbol{x}}_T = \boldsymbol{\mu}_x$.
- Estimated covariance matrix $\sigma^{-2}\boldsymbol{\Sigma}_x$.

The basic form of this algorithm underscores a greedy procedure that resembles the OMP introduced in [8]. Sparse reconstruction algorithms such as the SP [9] possess properties that produce far better results in terms of exact reconstruction frequency. Following the same rationale as in FMLM-$\boldsymbol{x}_i$, the inference procedure can improved by employing the key ideas behind the SP algorithm. The proposed algorithm termed Bayesian Subspace Pursuit is described below:

Algorithm 2 Bayesian Subspace Pursuit

Input: $\boldsymbol{\Phi}, \boldsymbol{y}, \sigma^2$
Initialise:
- Initialise iteration counter, $k = 1$.
- Calculate $\theta_i = |\boldsymbol{\phi}_i^T \boldsymbol{y}| - 1$ for $i \in [1, n]$.
- $T_k = \{i \in [1, n] : \theta_i > 0\}$.
- If $|T_k| = 0$ then $T_k = \{\text{index } i \in [1, n] \text{ for which } |\boldsymbol{\phi}_i^T \boldsymbol{y}| \text{ is maximised}\}$.
- Calculate $\alpha_i = \frac{1}{|\boldsymbol{\phi}_i^T \boldsymbol{y}|}$ for $i \in T_k$.

Iteration:
- Store $\alpha_{max} = \arg\max_{i \in T_k} |\alpha_i|$.
- Calculate values α_i and $\theta_i = q_i^2 - s_i$ for $i \in [1, n]$.
- Construct subsets: $T_{\theta_i > 0} = \{i \in [1, n] : \theta_i > 0\}$ and $T_{\alpha_i \leq a_{max}} = \{i \in [1, n] : |\alpha_i| \leq a_{max}, |\alpha_i| < +\infty\}$.
- If $|T_{\theta_i > 0}| = 0$ then $s = |T_{\alpha_i \leq a_{max}}| + 1$ else $s = |T_{\theta_i > 0}| + |T_{\alpha_i \leq a_{max}}|$.
- $T' = T_k \cup \{\text{indices corresponding to } s \text{ smallest values of } \alpha_i \text{ for } i \in [1, n]\}$.
- Increase counter, $k = k + 1$.
- Compute covariance matrix $\sigma^{-2}\boldsymbol{\Sigma}_x$ and $\boldsymbol{\mu}_x$.
- $T_k = \{\text{indices corresponding to } s \text{ largest non-zero values of } |\boldsymbol{\mu}_x| \text{ for which } \alpha_i > 0\}$
- If $\sigma^2 \mathcal{L}_k$ has reached a steady state then quit. Continue otherwise.

Output:
- Estimated support set T_k and sparse signal $\hat{\boldsymbol{x}}$ with $|T_k|$ non-zero components, $\hat{\boldsymbol{x}}_T = \boldsymbol{\mu}_x$.
- Estimated covariance matrix $\sigma^{-2}\boldsymbol{\Sigma}_x$.

IV. Empirical Results

In this section the performance of the proposed methods is assessed in simple simulated scenarios. Due to space constraints we only present results where the non-zero components of the signals take values from $\{+1, -1\}$ since this is a rather challenging case for OMP-like algorithms.

A. Single Time Instance

We concentrate on the performance of the algorithms for a single time instance and for $\sigma^2 = 0$. The algorithms under comparison are the FMLM algorithm as originally presented in [10], the FMLM-$\boldsymbol{x}_i$ described above and the BSP. In order to show the improvement over choosing basis vectors based on $\boldsymbol{x}_i$ we compare the results against another version, termed FMLM-$\boldsymbol{\alpha}_i$ which makes selections based on $\boldsymbol{\alpha}_i$. This version is not described here due to space constraints and since the differences with FMLM-$\boldsymbol{x}_i$ are very little. The experiment is as follows:

1) Generate matrix $\boldsymbol{\Phi} \in \mathbb{R}^{128 \times 256}$ with i.i.d entries drawn from $\mathcal{N}\left(0, \frac{1}{m}\right)$.
2) Generate support set T uniformly at random so that $|T| = K$.
3) Choose values for $\boldsymbol{x}_T$ from $\{+1, -1\}$ also at random.
4) Compute $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x}$ and apply the reconstruction algorithms mentioned above to get estimates $\hat{\boldsymbol{x}}$. Compare estimates to $\boldsymbol{x}$.
5) Repeat experiment for increasing values of K and for 100 realisations.

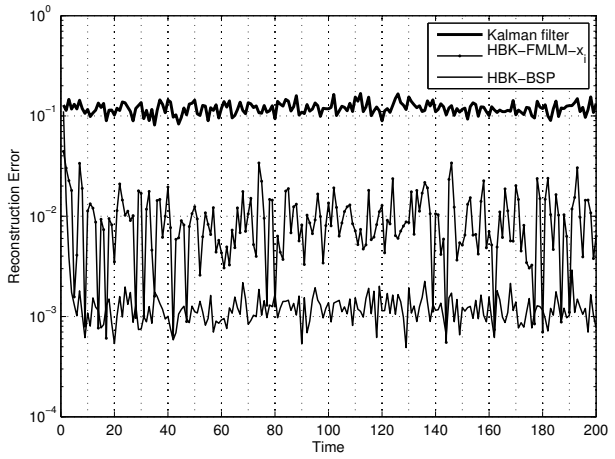

Fig. 2: Reconstruction error comparison for number of measurements, $m = 128$

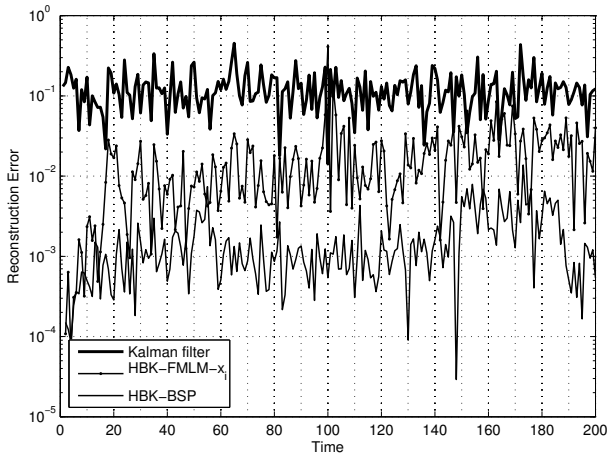

Fig. 3: Reconstruction error comparison for reduced number of measurements, $m = 6$

The results from this procedure are depicted in Figure 1. The first critical observation is that the original FMLM fails completely when $\sigma^2 = 0$ due to the improper scaling of the cost function by σ^2. By comparing FMLM-$\boldsymbol{x}_i$ and FMLM-$\boldsymbol{x}_i$ we observe the increase in exact reconstruction frequency by making selections based on $\boldsymbol{x}_i$. Finally the curve for the BSP algorithm shows that the employment of attributes from the SP makes the inference procedure substantially more efficient.

B. Dynamic Sparse Signal

We now compare the proposed method namely Hierarchical Bayesian Kalman filter (HBK) against the original Kalman filter. Signal $\boldsymbol{x}_t \in \mathbb{R}^n$ is chosen to be sparse in its natural basis. The indices i of the non-zero entries are chosen uniformly at random from $[1, n]$ where $n = 256$. 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 was $T = 200$ time instances. At two randomly chosen time instances $T = 50$ and $T = 150$ a change in the support of $\boldsymbol{x}_t$ is introduced. A non-zero component is added to the support of $\boldsymbol{x}_{50}$ and a non-zero component is removed from the support of $\boldsymbol{x}_{150}$. Apart from these two time instances the support of $\boldsymbol{x}_t$ remains stationary. At $T = 1$ the support is initialised with $K = 30$ non-zero components. Noise variance is set to $\sigma^2 = 0.01$ for the entire simulation time. We compare the following techniques; the classic Kalman filter, the HBK with FMLM-$\boldsymbol{x}_i$ as the optimisation procedure and the HBK with BSP.

Two scenarios are considered. In the first scenario noisy measurements $\boldsymbol{y}_t$ are taken by choosing the design matrix $\boldsymbol{\Phi}_t \in \mathbb{R}^{128\times 256}$ as described in subsection IV-A and is re-sampled at each time instance. The number of measurements m remains constant at each time instance. In Figure 2 we primarily observe how the HBK outperforms the original Kalman filter, direct consequence of the sparse dynamic model. We also notice that BSP outperforms FMLM-$\boldsymbol{x}_i$ since according to Figure 1 the latter is expected to fail for half of the times when $K = 30$.

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 $m = 6 < 2 \times 30$. This represents a difficult case since the number of measurements is less than twice the number of non-zero components which is 30 and tests the robustness of the proposed method. By observing Figure 3 we can see that HBK-BSP still performs optimally compared to the other two cases. Statistical information is retained for the BSP through the entire simulation time while it is gradually lost for the FMLM-$\boldsymbol{x}_i$ and reconstruction error increases.

REFERENCES

[1] N. Vaswani, "Kalman filtered compressed sensing," in *Image Processing, 2008. ICIP 2008. 15th IEEE International Conference on*, oct. 2008, pp. 893 –896.

[2] A. Carmi, P. Gurfil, and D. Kanevsky, "Methods for sparse signal recovery using kalman filtering with embedded pseudo-measurement norms and quasi-norms," *Signal Processing, IEEE Transactions on*, vol. 58, no. 4, pp. 2405 –2409, april 2010.

[3] J. Ziniel, L.C. Potter, and P. Schniter, "Tracking and smoothing of time-varying sparse signals via approximate belief propagation," in *Signals, Systems and Computers (ASILOMAR), 2010 Conference Record of the Forty Fourth Asilomar Conference on*. IEEE, 2010, pp. 808–812.

[4] A. Charles, M.S. Asif, J. Romberg, and C. Rozell, "Sparsity penalties in dynamical system estimation," in *Information Sciences and Systems (CISS), 2011 45th Annual Conference on*. IEEE, 2011, pp. 1–6.

[5] Michael E. Tipping, "The relevance vector machine," 2000.

[6] Shihao Ji, Ya Xue, and L. Carin, "Bayesian compressive sensing," *Signal Processing, IEEE Transactions on*, vol. 56, no. 6, pp. 2346 –2356, june 2008.

[7] D.P. Wipf and B.D. Rao, "Sparse bayesian learning for basis selection," *Signal Processing, IEEE Transactions on*, vol. 52, no. 8, pp. 2153 – 2164, aug. 2004.

[8] J.A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231 – 2242, oct. 2004.

[9] Wei Dai and Olgica Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Trans. Inform. Theory*, vol. 55, pp. 2230–2249, 2009.

[10] 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.