Forty-Eighth Annual Allerton Conference
Allerton House, UIUC, Illinois, USA
September 29 - October 1, 2010

# Bayesian Sequential Compressed Sensing in Sparse Dynamical Systems

Dino Sejdinović, Christophe Andrieu
School of Mathematics
University of Bristol
University Walk, Bristol BS8 1TW, UK
Email: {d.sejdinovic, c.andrieu}@bristol.ac.uk

Robert Piechocki
Department of Electrical & Electronic Engineering
University of Bristol
Merchant Venturers Building, Bristol, BS8 1UB, UK
Email: r.j.piechocki@bristol.ac.uk

***Abstract*—While the theory of compressed sensing provides means to reliably and efficiently acquire a sparse high-dimensional signal from a small number of its linear projections, sensing of dynamically changing sparse signals is still not well understood. We pursue a Bayesian approach to the problem of sequential compressed sensing and develop methods to recursively estimate the full posterior distribution of the signal.**

## I. Introduction

The advance of Compressed Sensing (CS) [8], [3], the theory underpinning the efficient and reliable reconstruction of a sparse signal from a limited number of its (possibly noisy) linear measurements, has received a significant research interest over the past few years. Let $\beta \in \mathbb{R}^N$ be an unknown vector which is sparse, i.e., at most $K \ll N$ of its entries are non-zero, and let:

$$y \quad = \quad X\beta + \epsilon, \tag{1}$$

where $X \in \mathbb{R}^{n \times N}$ is a known measurement matrix, $y \in \mathbb{R}^n$ is the observation and noise $\epsilon$ follows a zero-mean gaussian distribution $\epsilon \sim \mathcal{N}(0, \sigma^2 I_n)$. The results of CS have shown that with the judicious choice of $X$ (namely, for matrices that obey a restricted isometry property [3]), even when $n \ll N$, reliable reconstruction of $\beta$ is possible and computationally tractable. One of the common approaches to performing the signal reconstruction, referred to as Lasso [14], or Basis Pursuit Denoising (BPDN) [6], is given by:

$$\hat{\beta} = \arg\min_{\beta} \frac{1}{2}||y - X\beta||_2^2 + \lambda||\beta||_1, \ \lambda > 0, \tag{2}$$

for a suitable choice of regularization parameter $\lambda > 0$. This is a convex unconstrained optimization problem, which can be cast as a quadratic program.

Within the CS framework, one of the problems that has recently captured attention of various researchers [2], [1], [16] is that of the sequential estimation of dynamically changing sparse signals. For example, assume that the previously estimated sparse signal $\beta$ has undergone small changes, and that we observe a new batch of measurements

$$\breve{y} \quad = \quad \breve{X}\breve{\beta} + \breve{\epsilon}. \tag{3}$$

The aim is to recognise ways for our previous estimate of $\beta$ to aid reconstruction of the new signal $\breve{\beta}$, by evading the need to solve a new Lasso/BPDN problem, and thus to possibly reduce the number of necessary measurements whenever the signal undergoes a change.

This problem has been addressed by Asif and Romberg [2], who developed a suite of dynamic algorithms based on homotopy continuation principles. Similarly to the least angle regression (LARS) [10] for solving Lasso, these algorithms trace the piecewise linear path from the solution of the original optimization program (2) to the solution of the one based on a new batch of measurements in (3).

In addition, Vaswani [16] combines the established algorithms for CS with a Kalman filtering approach, by initially using CS to estimate the support of the signal, followed by a Kalman filter on the currently estimated support, which is corrected by additional runs of CS, whenever the filtering error increases. A similar approach was pursued in [1].

In this contribution, we pursue a Bayesian approach to sequential compressed sensing, and, in contrast to the previous work, propose a methodology for recursive estimation of the full posterior distribution of the signal at each time instance. Namely, a cloud of samples representing the posterior distribution is updated in a sequential way, following each new batch of measurements.

The next section overviews the Bayesian alternatives to point estimation via Lasso/BPDN, and demonstrates the use of Monte Carlo methods in tracking the posterior distribution. Section III outlines the state-space model for dynamically changing sparse signals, and establishes some of its properties, while Section IV proposes a sequential importance sampling with resampling to track sparse signal changes, and presents some empirical results. Section V concludes the paper.

## II. Bayesian approach to CS

### A. Lasso and sparsity promoting priors

Tibshirani [14] noticed that the Lasso estimate can be interpreted as a posterior mode estimate when the signal entries have independent and identical Laplace prior distributions. Namely, if we set a prior distribution of $\beta$ to:

$$p(\beta) = \prod_{j=1}^{N} \mathcal{L}(\beta_j;\lambda) = \prod_{j=1}^{N} \left(\frac{\lambda}{2} e^{-\lambda|\beta_j|}\right) \propto \exp\left(-\lambda||\beta||_1\right), \quad (4)$$

the posterior mode is simply:

$$\begin{aligned}\hat{\beta} &= \arg\max_{\beta} p(\beta|y) \\ &= \arg\max_{\beta} \frac{-1}{2\sigma^2}||y - X\beta||_2^2 + \log p(\beta) \\ &= \arg\min_{\beta} \frac{1}{2\sigma^2}||y - X\beta||_2^2 + \lambda||\beta||_1, \end{aligned} \quad (5)$$

which is exactly the Lasso estimate with the regularization parameter scaled by the noise variance. Many authors have used this connection, either to propose the use of other, similar, sparsity-promoting prior distributions [11] for point estimation, or to study Bayesian variants of the method [13], [12], aimed at describing $p(\beta|y)$ in other meaningful ways, such as confidence intervals, at the cost of being more computationally intensive. While such methods may not be practical in very high-dimensional settings, they can nonetheless be used to quantify uncertainty in estimating sparse signals.

The key observation from [13] is that the Laplace distribution can be written as a scale mixture of zero-mean gaussian distributions with an exponential mixing density:

$$\mathcal{L}(\beta_j;\lambda) = \int_0^{\infty} \mathcal{N}(\beta_j; 0, \theta_j) g_\lambda(\theta_j) d\theta_j, \quad (6)$$

where

$$g_\lambda(\theta_j) = \frac{\lambda^2}{2} \exp\left(-\frac{\lambda^2}{2}\theta_j\right). \quad (7)$$

This means that after conditioning on the hyperparameters $\theta_1, \theta_2, \ldots, \theta_N$, $\beta$ is simply a multivariate gaussian:

$$\beta|\theta_1, \theta_2, \ldots, \theta_N \sim \mathcal{N}(0_N, D(\theta)), \quad (8)$$

where $D(\theta) = \text{diag}(\theta_1, \theta_2, \ldots, \theta_N)$.

It turns out that this representation as a scale mixture of gaussians is shared by many other sparsity-promoting distributions studied in the literature [5], [13]. For example, one such model [4] considers generalized hyperbolic distributions and assigns an independent and identical scale mixture of gaussians to each signal entry $\beta_j$, $j \in \{1, 2, \ldots, N\}$:

$$p(\beta_j) = \int_0^{\infty} \mathcal{N}(\beta_j; 0, \theta_j) p(\theta_j) d\theta_j, \quad (9)$$

where $\theta_j$, $j \in \{1, 2, \ldots, N\}$, follows a generalized inverse gaussian distribution $\mathcal{GIG}(\nu, \delta, \lambda)$ with density

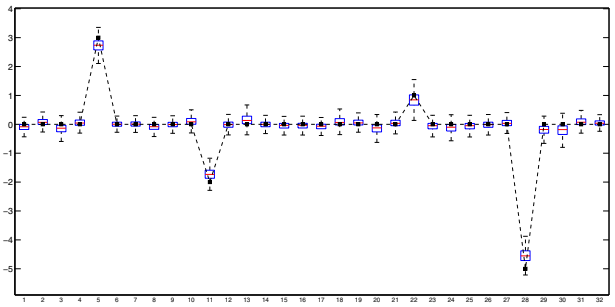

Figure 1. Posterior distribution $p(\beta|y)$ estimated by Gibbs sampler

$$\mathcal{GIG}(\theta_j;\nu,\delta,\lambda) = \frac{(\lambda/\delta)^\nu}{2K_\nu(\lambda\delta)} \theta_j^{\nu-1} \exp\left[-\frac{1}{2}(\delta^2\theta_j^{-1} + \lambda^2\theta_j)\right], \quad (10)$$

where $K_\nu(z)$ is the modified Bessel function of the second kind.

For $\nu = 1$, $\delta = 0$, this model recovers a Laplace prior distribution, but it encompasses others, including the normal gamma law and the normal inverse gaussian law [5].

### B. Sampling from $p(\beta|y)$

Assuming a general prior in (9), we can easily sample from $p(\beta|y)$ by sampling from $p(\beta, \theta|y)$ using the Gibbs sampler, based on the following full conditionals: for $j = \{1, 2, \ldots, N\}$,

$$\theta_j|\beta_j, y \sim \mathcal{GIG}\left(\nu - 1/2, \sqrt{\delta^2 + \beta_j^2}, \lambda\right), \quad (11)$$

and

$$\beta|\theta, y \sim \mathcal{N}\left(\mu^{(i)}, \Sigma^{(i)}\right), \quad (12)$$

where

$$\begin{aligned}\mu^{(i)} &= \left(X^\top X + \sigma^2 D(1/\theta^{(i)})\right)^{-1} X^\top y \\ \Sigma^{(i)} &= \sigma^2 \left(X^\top X + \sigma^2 D(1/\theta^{(i)})\right)^{-1}. \end{aligned} \quad (13)$$

The conditional (11) is obtained by noticing that $p(\theta|\beta, y) \propto p(y|\beta) \prod_{j=1}^{N} p(\beta_j|\theta_j) p(\theta_j)$, while (12) is a simple Bayesian linear regression. Thus, both conditional distributions are available in closed form and can be sampled from exactly. The details on how to efficiently sample from generalized inverse gaussian distribution are available in [7].

In Fig. 1, a boxplot of a posterior distribution $p(\beta|y)$ estimated by a Gibbs sampler for example where $N = 32$, $n = 16$, $K = 4$, $\sigma^2 = 0.05$, $\nu = 1$, $\delta = 0$ and $\lambda = 0.05$. Black squares denote the true value of the signal - they are within the credible intervals for each of the entry. We have initialized Gibbs sampler with $\beta^{(0)} = X^\top y$.

## III. Sparse dynamical system - model and properties

### A. Model

We will consider the following state-space model:

$$\begin{aligned} \beta^t &= P_t\beta^{t-1} + z^t \\ y^t &= X_t\beta^t + \epsilon^t, \end{aligned} \tag{14}$$

where for all $t$, $P_t$ is a known orthogonal $N \times N$ matrix, $z^t$ is the innovation in the signal evolution, $X_t$ is a known $n \times N$ measurement matrix and the entries of $\epsilon^t$ follow a zero mean gaussian distribution of known variance $\sigma^2$. Since we would like our model to capture the evolution of the sparse signals $\beta^1, \beta^2, \ldots, \in \mathbb{R}^N$ which share "common sparsity", we can think of $P_t$ as being either the identity or a known permutation matrix, but the model properties hold regardless of this simplification. In addition, we assume that $z^t$ is also sparse, and thus we aim to simultaneously enforce the following prior information for all $t$ and for $j \in \{1, 2, \ldots, N\}$,

$$\begin{aligned} p(\beta_j^t) &= \int_0^\infty \mathcal{N}(u_j^t; 0, u_j^t)p(u_j^t)du_j^t, \\ p(z_j^t) &= \int_0^\infty \mathcal{N}(z_j^t; 0, v_j^t)p(v_j^t)dv_j^t, \end{aligned} \tag{15}$$

where $u_j^t$ and $v_j^t$ follow the identical and independent generalized inverse gaussian distributions $\mathcal{GIG}(\nu, \delta, \lambda_u)$ and $\mathcal{GIG}(\nu, \delta, \lambda_v)$ respectively. This model combines two kinds of prior information: temporal dependence and sparsity of the signal at each time step. It can be easily shown that, in the case of Laplacian priors, the mode of the posterior $p(\beta^t|y, \beta^{t-1})$ is given by:

$$\begin{aligned} \hat{\beta}^t = \arg\min_\beta \frac{1}{2}||y^t - X_t\beta||_2^2 + \\ + \lambda_u||\beta||_1 + \lambda_v||\beta - P_t\beta^{t-1}||_1,\ \lambda_u, \lambda_v > 0 \end{aligned} \tag{16}$$

which is a point estimate similar to that of the *fused lasso* by Tibshirani *et al* [15].

### B. Conditionally linear gaussian structure

In the rest of this section we will use the following proposition.

**Proposition 1.** *Let $Q$, $R$ be $N \times N$ covariance matrices, and let $B$ be an $N \times N$ matrix. Then, $\forall \beta, \breve{\beta} \in \mathbb{R}^N$,*

$$\begin{aligned} \mathcal{N}(\breve{\beta}; \mathbf{0}, Q)\mathcal{N}(\beta; B\breve{\beta}, R) = \\ \mathcal{N}(\breve{\beta}; \Sigma B^\top R^{-1}\beta, \Sigma)\mathcal{N}(\beta; \mathbf{0}, R + BQB^\top) \end{aligned} \tag{17}$$

*where $\Sigma = (Q^{-1} + B^\top R^{-1}B)^{-1}$.*

*Proof:* We start with

$$\begin{aligned} \mathcal{N}(\breve{\beta}; 0, Q)\mathcal{N}(\beta; B\breve{\beta}, R) = \\ \frac{1}{(2\pi)^N\sqrt{\det(Q)\det(R)}}\exp(-\frac{1}{2}\Phi), \end{aligned} \tag{18}$$

where

$$\begin{aligned} \Phi = \breve{\beta}^\top Q^{-1}\breve{\beta} + (\beta - B\breve{\beta})^\top R^{-1}(\beta - B\breve{\beta}) = \\ \breve{\beta}^\top(Q^{-1} + B^\top R^{-1}B)\breve{\beta}^\top - 2\breve{\beta}^\top B^\top R^{-1}\beta + \beta^\top R^{-1}\beta. \end{aligned} \tag{19}$$

By substituting

$$\begin{aligned} \Sigma &= (Q^{-1} + B^\top R^{-1}B)^{-1} \\ \mu &= (Q^{-1} + B^\top R^{-1}B)^{-1}B^T R^{-1}\beta, \end{aligned} \tag{20}$$

we obtain:

$$\begin{aligned} \Phi &= (\breve{\beta} - \mu)^\top\Sigma^{-1}(\breve{\beta} - \mu) + \\ &+ \beta^\top(R^{-1} - R^{-1}B(Q^{-1} + B^\top R^{-1}B)^{-1}B^\top R^{-1})\beta \\ &= (\breve{\beta} - \mu)^\top\Sigma^{-1}(\breve{\beta} - \mu) + \beta^\top(R + BQB^\top)\beta, \end{aligned} \tag{21}$$

where in the last step we used the *Woodbury matrix identity*. Also,

$$\begin{aligned} \det\left((Q^{-1} + B^\top R^{-1}B)^{-1}\right) \cdot \det(R + BQB^\top) = \\ \frac{\det(R) \cdot \det(I + B^\top R^{-1}QB)}{\det(Q^{-1}) \cdot \det(I + BR^{-1}QB^\top)} = \\ \det(R) \cdot \det(Q), \end{aligned} \tag{22}$$

which together with (21), proves the claim. In the first step of (22), we used $\det(M + XY) = \det(M)\det(I + YM^{-1}X)$ for any invertible matrix $M$, while in the second step we used that $\det(M^\top) = \det(M)$. ∎

Now, since

$$\begin{aligned} p(\beta^t|\beta^{t-1}, u^t, v^t) \propto \\ p(\beta^t|u^t)p(\beta^{t-1}|\beta^t, v^t) \propto \\ \mathcal{N}(\beta^t; 0, D(u^t))\mathcal{N}(\beta^{t-1}; P_t^\top\beta^t, P_t^\top D(v^t)P_t), \end{aligned} \tag{23}$$

by applying Proposition 1 to the above product of gaussian densities, we see that the conditional probability $p(\beta^t|\beta^{t-1})$ can be written as a mixture of multivariate gaussian distributions:

$$\begin{aligned} p(\beta^t|\beta^{t-1}) \propto \int\int \mathcal{N}\left(\beta^t; D(\frac{u^t v^t}{u^t + v^t})P_t\beta^{t-1}, D(\frac{u^t v^t}{u^t + v^t})\right) \cdot \\ g\left(u^t, v^t|\beta^{t-1}\right) du^t dv^t, \end{aligned} \tag{24}$$

where the mixing density is given by

$$g\left(u^t, v^t|\beta^{t-1}\right) \propto \prod_{j=1}^{N}\left[\mathcal{GIG}(u_j^t;\nu,\delta,\lambda_u)\mathcal{GIG}(v_j^t;\nu,\delta,\lambda_v)\right]\cdot \mathcal{N}(\beta^{t-1};0,P_t^\top D(u^t+v^t)P_t) \quad (25)$$

This can be interpreted in the following way: given some hyperparameters $u^t$ and $v^t$, our model is equivalent to the following linear gaussian dynamical system:

$$\begin{aligned}\beta^t &= D(\frac{u^t}{u^t+v^t})P_t\beta^{t-1}+\mathcal{N}\left(0,D(\frac{u^tv^t}{u^t+v^t})\right),\\ y^t &= X_t\beta^t+\mathcal{N}(0,\sigma^2 I_n), \end{aligned} \quad (26)$$

which can be solved exactly by a Kalman filter. Note that $D(\frac{u^t}{u^t+v^t})P_t$ in (26) can be interpreted as a "shrinkage" matrix, which effectively plays the role of keeping the evolving signal $\beta^t$ approximately sparse.

## IV. Sequential Monte Carlo sampler

We will now investigate a method based on sequential importance sampling with resampling (see, e.g., [9] and references therein), which aims to represent the full posterior distribution $p(\beta^{0:T}|y^{1:T})$ by sampling from $p(\beta^{0:T},u^{1:T},v^{1:T}|y^{1:T})$. The main idea of the method is to track a cloud of samples that represents $p(\beta^t|y^{1:t})$ and update it in a sequential way. Based on the underlying conditionally linear gaussian structure of the state-space model, at every time step, given a cloud of samples of $u^t$,$v^t$ and $\beta^{t-1}$, we can find the exact corresponding distributions of $\beta^t$ using the Kalman Filtering equations. Note that, if no estimate of $\beta^0$ is available, samples from $p(\beta^1|y^1)$ can be obtained via the Gibbs sampler described in Section 2. This suggests a sequential Monte Carlo sampler which uses importance sampling with resampling to obtain a cloud of samples of $u^t$, $v^t$ first, and then samples from the conditional posterior $p(\beta^t|\beta^{t-1},y^t,u^t,v^t)$, which is just a multivariate gaussian. Thus, the overall sampling distribution is:

$$\pi(\beta^{0:T},u^{1:T},v^{1:T}|y^{1:T}) \propto \prod_{t=1}^{T}\Bigg[p(\beta^t|\beta^{t-1},y^t,u^t,v^t)\cdot \pi(u^t,v^t|\beta^{0:t-1},y^{1:t})\Bigg]. \quad (27)$$

At the $t$-th time step, given $S$ samples $\left\{\beta^{0:t,(i)}\right\}_{i=1}^{S}$ from (27), we approximate the posterior distribution $p(\beta^{0:t}|y^{1:t})$ by:

$$\hat{p}(\beta^{0:t}|y^{1:t}) = \sum_{i=1}^{S} w_t^{(i)}\delta(\beta^{0:t,(i)}). \quad (28)$$

The importance weights $w_t^{(i)}$ in the implementation of the sequential importance sampling with resampling can be evaluated recursively as successive observations become available as:

$$w_t^{(i)} \propto w_{t-1}^{(i)}\cdot \frac{p(y^t|u^{t,(i)},v^{t,(i)},\beta^{t-1,(i)})g(u^{t,(i)},v^{t,(i)}|\beta^{t-1,(i)})}{\pi(u^t,v^t|\beta^{0:t-1},y^{1:t})}, \quad (29)$$

where $p(y^t|u^t,v^t,\beta^{t-1})$ is a multivariate gaussian, given by:

$$p(y^t|u^t,v^t,\beta^{t-1}) = \mathcal{N}\left(y^t;X_tD\left(\frac{u^t}{u^t+v^t}\right)P_t\beta^{t-1},X_tD\left(\frac{u^tv^t}{u^t+v^t}\right)X_t^\top\right). \quad (30)$$

If the effective sample size $S_{eff} = 1/\sum_{i=1}^{S}\left(w_t^{(i)}\right)^2$ is smaller than a prescriped threshold $\widetilde{S}$, we perform the resampling step. Following the resampling, we obtain the sample $\beta^{t,(i)}$ from

$$p(\beta^{t,(i)}|\beta^{t-1,(i)},y^t,u^{t,(i)},v^{t,(i)}) = \mathcal{N}(\beta^{t,(i)};\mu^{t,(i)},\Sigma^{t,(i)}), \quad (31)$$

where

$$\begin{aligned}\mu^{t,(i)} &= \left(\frac{1}{\sigma^2}X_t^\top X_t + D\left(\frac{u^{t,(i)}+v^{t,(i)}}{u^{t,(i)}v^{t,(i)}}\right)\right)^{-1}\cdot \left(D\left(\frac{1}{v^{t,(i)}}\right)P_t\beta^{t-1,(i)}+\frac{1}{\sigma^2}X_t^\top y^t\right),\\ \Sigma^{t,(i)} &= \left(\frac{1}{\sigma^2}X_t^\top X_t + D\left(\frac{u^{t,(i)}+v^{t,(i)}}{u^{t,(i)}v^{t,(i)}}\right)\right)^{-1}. \end{aligned} \quad (32)$$

There are various different ways to make a judicious choice of importance distribution $\pi(u^t,v^t|\beta^{0:t-1},y^{1:t})$ in (27), but, for clarity of exposition, we will focus on a simple case where $\pi(u^t,v^t|\beta^{0:t-1},y^{1:t}) = g\left(u^t,v^t|\beta^{t-1}\right)$. This choice is very convenient, as in the case where $\lambda_u=\lambda_v=\lambda$, and $\delta=0$, the entries of the sum of $u_t$ and $v_t$ can be shown to follow a generalized inverse gaussian distribution:

$$p(u_j^t+v_j^t=\theta_j^t|\beta_j^{t-1}) = \mathcal{GIG}\left(\theta_j^t;\nu+1/2,\beta_j^{t-1},\lambda\right). \quad (33)$$

This means that we can sample exactly and easily from $g\left(u^t,v^t|\beta^{t-1}\right)$, since we have

$$p(u_j^t|\theta_j^t) = \frac{\mathbf{1}\{0\le u_j^t\le\theta_j^t\}}{\theta_j^t}. \quad (34)$$

In Figures 2-5, we have presented some preliminary results of applying a sequential Monte Carlo sampler to the dynamically evolving sparse signals, showing the boxplots of the estimated posterior distributions and representing both the case where the signal does not undergo any support change (Figures 2 and 3) and the case where two changes in the support occur

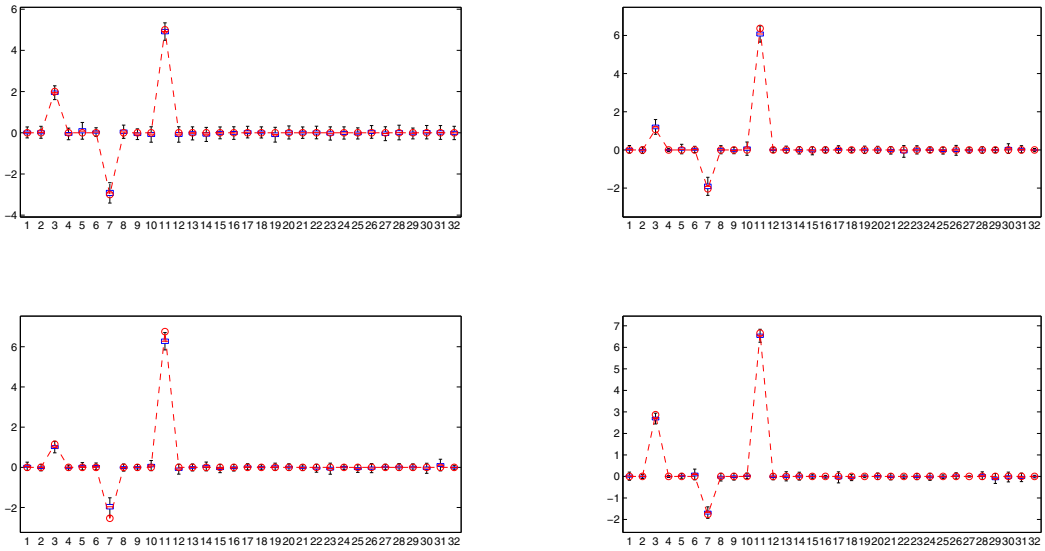

Figure 2. The iterations 1-4 of the sequential Monte Carlo sampler, no change in support

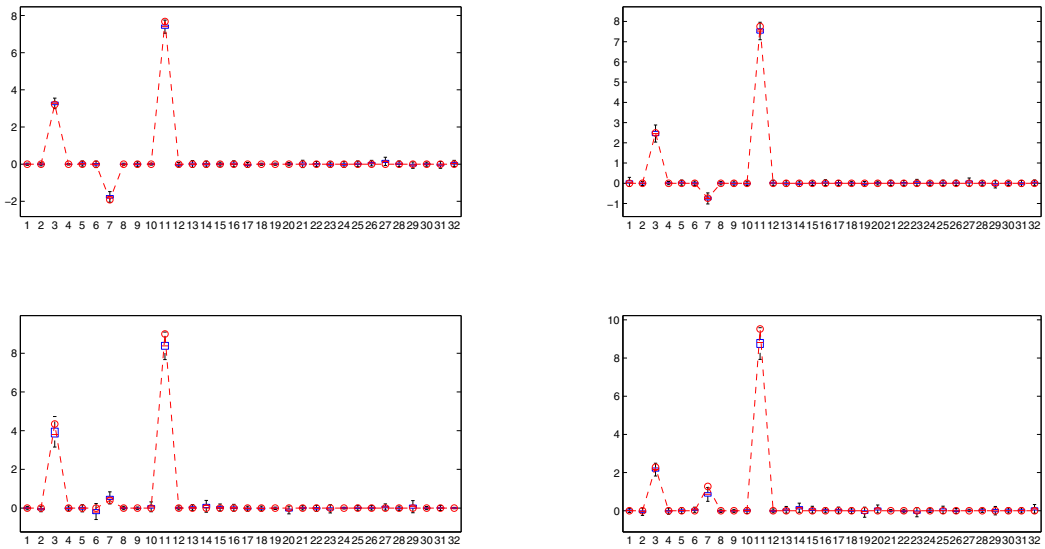

Figure 3. The iterations 5-8 of the sequential Monte Carlo sampler, no change in support

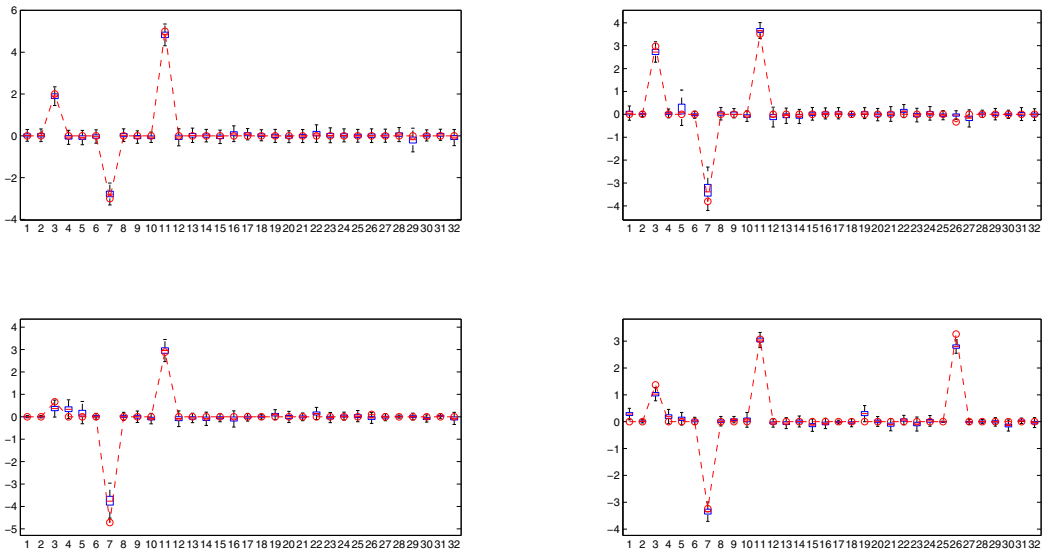

Figure 4. The iterations 1-4 of the sequential Monte Carlo sampler, change in support

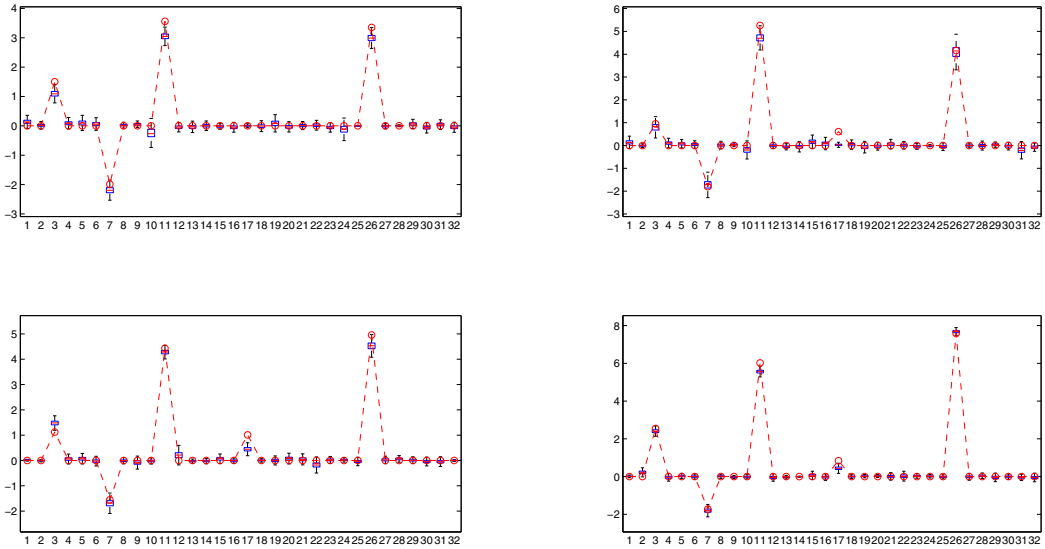

Figure 5. The iterations 5-8 of the sequential Monte Carlo sampler, change in support

(Figures 4 and 5). The parameters were $N = 32$, $n = 10$, $\sigma^2 = 0.05$, $\nu = 1$, $\delta = 0$ and $\lambda = 0.05$. We observe that the true signal values marked in open circles lie in most cases within the credible intervals, indicating that the method successfully updates the posterior distribution based on the new batch of measurements. While in many cases the posterior distribution appears peaked, implying that a point estimate based on either the mode or the median of the posterior would be sufficient, instances can also be observed where the credible interval is much wider, illustrating the utility of the method in quantifying uncertainty in the signal reconstruction.

## V. Conclusions

We derive a Bayesian framework for compressed sensing of dynamically changing sparse signals and present a hierarchical model for performing sequential estimation of the full posterior distribution. We propose a method of tracking posterior distributions recursively via sequential Monte Carlo sampling, and show how our model of a sparse dynamical system can be represented as a conditionally linear Gaussian state space model, leading to some interesting analytical properties of the method, as many quantities involved can be calculated by using Kalman filtering equations. A more judicious choice of importance functions may improve the efficiency of the method, by allowing us to track smaller sample sizes. For example, in the case of the Laplace priors, the fact that the posterior modes can be estimated successively by convex optimization may be utilized in the design of importance functions.

## Acknowledgment

The authors would like to thank Olgica Milenković for insightful discussions related to this work.

## References

[1] D. Angelosante, E. E. Grossi, and G.B. Giannakis. Compressed sensing of time-varying signals. In *Proc. of The 16th International Conference on Digital Signal Processing*, Santorini, Greece, 2009.

[2] M.S. Asif and J. Romberg. Dynamic updating for l1 minimization. *IEEE Journal of Selected Topics in Signal Processing*, 4(2):421 –434, apr. 2010.

[3] E. J. Candes, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59:1207–1223, 2006.

[4] F. Caron. Hierarchical models for dependent sparse linear regressions. June 2010. presented at Workshop on sparse structures: statistical theory and practice, Bristol, UK.

[5] F. Caron and A. Doucet. Sparse Bayesian nonparametric regression. In *Proc. of the 25th International Conference on Machine Learning (ICML)*, pages 88–95, 2008.

[6] S. Chen, D. Donoho, and M. Saunders. Atomic decomposition by basis pursuit. *SIAM Jour. Sci. Comp.*, 20(1):33–61, 1998.

[7] J.S. Dagpunar. An easily implemented generalized inverse gaussian generator. *Commun. Statist. Simul.*, 18(2):703–710, 1989.

[8] D. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[9] A. Doucet, S.J. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10:197–208, 2000.

[10] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Ann. Statist.*, 32(2):407–499, 2004.

[11] M. Figueiredo. Adaptive sparseness for supervised learning. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 25:1150–1159, 2003.

[12] Shihao Ji, Ya Xue, and L. Carin. Bayesian compressive sensing. *IEEE Transactions on Signal Processing*, 56(6):2346–2356, June 2008.

[13] T. Park and G. Casella. The Bayesian Lasso. *Journal of the American Statistical Association*, 103:681–686, 2008.

[14] R. Tibshirani. Regression shrinkage and selection via the lasso. *J. Royal Statist. Soc. B*, 58:267–288, 1996.

[15] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight. Sparsity and smoothness via the fused lasso. *J. Royal Statist. Soc. B*, 67:91–108, 2005.

[16] N. Vaswani. Kalman filtered compressed sensing. In *Proc. of the 15th IEEE International Conference on Image Processing (ICIP).*, pages 893 –896, oct. 2008.