Sample Size Reduction in Groundwater Surveys Via Sparse Data Assimilation

Zaeem Hussain
Department of Computer Science
LUMS School of Science and Engineering
Lahore, Pakistan
Email: zaeem.hussain@gmail.com

Abubakr Muhammad
Department of Electrical Engineering
LUMS School of Science and Engineering
Lahore, Pakistan
Email: abubakr@lums.edu.pk

***Abstract*—In this paper, we focus on sparse signal recovery methods for data assimilation in groundwater models. The objective of this work is to exploit the commonly understood spatial sparsity in hydrodynamic models and thereby reduce the number of measurements to image a dynamic groundwater profile. To achieve this we employ a Bayesian compressive sensing framework that lets us adaptively select the next measurement to reduce the estimation error. An extension to the Bayesian compressive sensing framework is also proposed which incorporates the additional model information to estimate system states from even lesser measurements. Instead of using cumulative imaging-like measurements, such as those used in standard compressive sensing, we use sparse binary matrices. This choice of measurements can be interpreted as randomly sampling only a small subset of dug wells at each time step, instead of sampling the entire grid. Therefore, this framework offers groundwater surveyors a significant reduction in surveying effort without compromising the quality of the survey.**

I. Introduction

The focus of this paper is to investigate techniques for state estimation using a small number of measurements and to propose a framework for data assimilation using those techniques, with a special emphasis groundwater monitoring. For many geophysical processes, groundwater being a case in point, the number of available measurements is limited. Data collection for groundwater models is especially cumbersome because the measurement process for groundwater level, flow or contamination involves digging of wells, collection of water samples and subsequent chemical analysis. Digging wells for the entire field may not be an economically justified option. Even if the wells are already dug (as is the case for agricultural lands in many provinces of Pakistan and India), collecting data still is an expensive and time consuming exercise. Hence a data assimilation technique that makes the most of the limited data available is most desirable. Another desirable feature of such a technique would be the adaptive nature of its measurements so that we can theoretically predict which observations points are better than others to conduct the next round of survey such that the new observations would reduce the estimation error the most.

Recently, techniques have been introduced for signal reconstruction that exploit a sparse representation in some transform domain. Most of such techniques have been inspired by the *Compressive Sensing* (CS) paradigm. However, these techniques either do not explicitly incorporate model dynamics or make additional assumptions about the system. In one of the earliest examples of such considerations, the work by Jafarpour et al [6] focuses specifically on hydrogeological reservoir systems. It uses a Discrete Cosine Transform (DCT) parametrization for states as such systems naturally exhibit a sparse representation in the spatial domain. They use the l_1 norm optimization technique to promote sparse signal recovery and performs data assimilation by minimizing the l_2 norm of the observations with the estimated states. History matching is done by placing weights and regularization parameters on each of the terms and the collective cost function thus obtained is minimized to estimate the system state vector. However, there is no procedure defined for setting or adjusting regularization coefficients and weights and the paper mentions that their value can be set from prior belief. Another example of using compressive sensing like ideas is [13] which studies the application of the compressive sensing procedure for soil moisture levels.

The framework introduced in [10] also exploits the sparsity of a given dynamical system in some transform domain and merges it with compressed sensing framework. It works by defining a support set of the system state as the indices at which the transform of the state vector is non zero. This framework then runs a reduced order Kalman filter for the non zero elements of the state transform coefficients corresponding to the support set. As the support set changes over time, this change is detected by the error between the estimate and the observations. The change is then estimated by performing the l_1 norm minimization of the vector that represents the supports set, based on the assumption that the sparsity pattern of the state vector changes slowly over time and thus the vector representing change in the non zero indices will be sparse. Thus this framework makes the additional assumption about the system that the sparsity pattern of the system states changes slowly with time. Moreover, it also requires knowledge of the support set for the initial state.

This work differs from all previous approaches in two respects. First, we have tackled the issue of point measurements instead of imaging-like cumulative measurements used in conventional compressive sensing. This opens the avenue for many other geophysical applications where sweep measurements are not available but where sensors probe the field directly at sample points. Secondly, we have given a Bayesian framework that tracks not just point estimates (as in the case of l_1-optimization in CS) but also their probability distribution. This enables one to study adaptation (between and during field measurements) as well as a formal way to incorporate sparse measurements into a Bayesian filtering framework. Although the later has not been been fully developed in this paper, glimpses of both adaptation and model incorporation have been demonstrated on two examples via simulation results. In both examples, one a toy 1D-heat equation and the other a 2D contamination spread model in groundwater, we have exploited the diffusive nature of the dynamics and demonstrated adaptation to changing sparsity and model incorporation.

II. Problem Formulation

We now formulate the problem following the standard data assimilation terminology [5]. Given a domain D with boundary ∂D parameterized by coordinates $x \in \mathbb{R}^d$, the spatio-temporal dynamics of a groundwater-like geophysical model are given by

$$\frac{\partial \psi(x,t)}{\partial t} = G(\psi(x,t), \alpha(x)) + u(x,t). \quad (1)$$

Here, the field to be estimated at each time step is $\psi(x,t) \in \mathbb{R}^N$, where N is the number of model state variables describing the field, each variable dependent on time and space. $G(\psi, \alpha)$ is the nonlinear model operator, $\alpha(x)$ are spatially varying parameters and u is a deterministic control input or a noisy forcing function. Appropriate initial and boundary conditions are also needed to solve the model. In most data assimilation problems, parameters $\alpha(x)$ are also unknown or at best poorly known. In groundwater flow models these parameters correspond to ground porosities or other geological structures. Here, we do not use this distinction between state and parameters and augment the state if such a situation arises.

The observation process is defined generically as a measurement functional $\mathcal{M}[\psi]$. In this paper, we focus on measurement functionals of the form

$$\mathcal{M}_i[\psi] = \int\int \psi^T(x,t)\delta_{\psi_i}\delta(t-t_i)\delta(x-x_i)dtdx, \quad (2)$$

which model point-sampling the field at location x_i and time t_i). To simply notation, we will denote such measurements as an M-dimensional vector $y \in \mathbb{R}^M$. Assumptions on sensor noise can also be built into this simplified model.

We make one additional assumption on this standard data assimilation framework. In many geophysical problems and in particular due to the diffusive properties of fluid flows, it is reasonable to assume that the field to be estimated will not have very sharp spatial variations. This can be technically captured by saying that for a fixed time, the field $\psi(x,.)$ is *spatially sparse* in a known transform domain. Specifically, if we take a Fourier-like transform of $\psi(x,.)$, then the field can be reconstructed using the knowledge of only a fraction of the transform coefficients.

We now state the problem as follows. *What is the minimum number of measurements M to estimate the joint pdf of parameters and model state, given a dynamical model with known uncertainties and that for a fixed time, the field $\psi(x,.)$ is spatially sparse?*

III. Reconstructing Sparse Fields

The compressive sensing paradigm works by exploiting the prior knowledge that the signal to be recovered is sparse or compressible in some transform domain. Consider an $N \times 1$ column vector $\mathbf{x}$ in R^N. The signal $\mathbf{x}$ can be represented in another basis of $N \times 1$ vectors $\{\mathcal{B}_i\}_{i=1}^N$ as

$$\mathbf{x} = \sum_{i=1}^{N} s_i \mathcal{B}_i, \qquad \text{or} \qquad \mathbf{x} = \mathcal{B}s, \quad (3)$$

where s is the column vector of the weights $s_i = \mathcal{B}_i^T\mathbf{x}$, and $\mathcal{B}$ is the $N \times N$ transformation matrix with $\mathcal{B}_i$ as its columns. Let y be the $M \times 1$ column vector of measurements given by

$$y = \Phi\mathbf{x} = \Phi\mathcal{B}s = \boldsymbol{\theta}s, \quad (4)$$

where Φ is the $M \times N$ measurement matrix. In our setup, $\mathbf{x}$ can be thought of as a grid representation of a static field $\psi(x,.)$ and Φ is a matrix model for the measurement functional $\mathcal{M}$ introduced in Equation (2). A good choice of transform $\mathcal{B}$ to model spatial sparsity is the DCT. A vector representation of a signal s is called K-sparse if it has only K non-zero coefficients. This is measured by the l_0 norm of the signal $\|s\|_0$.

A. Standard Compressive Sensing Framework

The general problem of reconstructing $\mathbf{x}$ from y with the measurement matrix Φ is equivalent to the problem of recovering s from y with the transformed $M \times N$ measurement matrix $\boldsymbol{\theta}$. Thus to perform reconstruction using the compressive sensing framework, we choose a transform domain in which s has a highly sparse representation. The problem of recovering $N \times 1$ column vector s from the $M << N$ measurements in y is in general under-determined and a unique solution to the problem does not exist. However, if the signal to be recovered is sparse, and the indices of the non zero elements in the signal are known, then recovery is possible if the number of measurements is greater than the number of non zero elements in the signal to be recovered. A necessary and sufficient condition for this to hold is that multiplication of $\boldsymbol{\theta}$ with any vector v that has the same non zero elements as s should not alter the length of v:

$$1 - \epsilon \leq \frac{||\boldsymbol{\theta}v||_2}{||v||_2} \leq 1 + \epsilon,$$

for a small $\epsilon > 0$. If the indices of the non zero elements in the vector to be recovered are not known, then a sufficient condition for recovery is that (5) should hold for any arbitrary vector with three times the number of non zero elements as in the original signal. This is known as the Restricted Isometric property [3], [4]. Another requirement for recovery of the signal is that the columns of $\mathcal{B}$ should not be combinations of a small number of rows of Φ and also the other way round. In other words, the two matrices should be incoherent and this condition is referred to as *incoherence* [3], [4].

To construct a measurement matrix which satisfies the above conditions would mean checking (5) for all possible arrangements of the non zero elements in the R^N vector space. However, we can simply select Φ as a random matrix whose entries are chosen from independent and identically distributed Gaussian variables [3], [4]. demonstrate that such a matrix would satisfy both the *RIP* and the incoherence property with high probability. The recovery algorithm itself is usually via an optimization program that minimizes the $\|s\|_0$ while obeying the basic measurement equation (4). The breakthrough provided by CS framework is that this difficult combinatorial problem can be solved by a relaxation using the l_1 norm instead [2]. The success of the CS framework also lies in the fact that such a recovery is possible with M of the order of $K\log(N/K)$, for a K-sparse signal which is an astonishingly low number compared to measuring $\mathbf{x}$ directly using N measurements.

B. CS with Sparse Measurement Matrices

In the standard framework described above, typical CS measurements created from Gaussian random matrices are themselves dense. This means that each measurement is a cumulative or weighted average of *all* the states variables. This goes against the requirement of point measurements that are themselves often very sparse for groundwater monitoring, as explained above. Thus the measurement matrix itself needs to be sparse, in addition to satisfying RIP and incoherence properties.

The construction of sparse matrices can be done in multiple ways. Although, we are forced to place all zeros in the columns for which the data points are unavailable, we are free to place non-zero values in the rest of the columns in any way we like. For example, we can randomly place a certain number of ones in each of these columns at a random row numbers. We can also place a single one in each of these columns in a different row number for every column. This would mean there is only a single one in each row and the number of rows will be equal to the number of elements of x that are available for observation. Such a matrix is more sparse than the one which randomly places ones in the *free* columns but is certainly more desirable for storage purposes. Our simulation results have shown that, for the missing data case, it does not make a difference if the binary matrix contains multiple ones in the columns for which the data is available or if it contains a single one in a different row in each column. We borrow this construction from [1]. Thus, in our simulation and experiments, we use sparse binary matrices for measurement which have as many rows as the number of points whose observation is available and contain a single one in each row for each of the available data points. It is worth noting here that the Binary matrices constructed as defined above have also been used in [13] which studies the application of the compressive sensing procedure for soil moisture levels. Although, the paper uses the binary matrices constructed in this way with a different transformation basis, the reconstruction results that they have reported indicate that such matrices should work well in practice with the right amount of available data.

C. Bayesian Compressive Sensing

The sparse signal recovery problem discussed earlier can also be viewed from a Bayesian viewpoint. The framework is inspired by work on Relevance Vector machines (RVM) [9] and provides the probability distributions for the unknowns instead of the point estimates provided by l_1 norm based optimization techniques. Thus, this framework provides additional information about the unknown signal and the resultant probability distributions can also be used for further analysis. It starts from the same linear regression problem

$$y = \boldsymbol{\theta} s + n, \tag{5}$$

where n is additive sensor noise. The RVM works by assuming a prior distribution on s for a sparse estimate and gives the distributions for the non zero weights in s.

Assuming the additive noise n to be zero mean Gaussian with variance σ^2, the distribution for y can be written as [7]

$$p(y|s,\sigma^2) = (2\pi\sigma^2)^{-K/2} exp(-\frac{1}{2\sigma^2}||y-\boldsymbol{\theta} s||^2), \tag{6}$$

where K, as before, counts the number of non zero elements in s or its sparsity. This is the Gaussian distribution for the training data y using the Bayesian approach.

Now we introduce *prior* knowledge that s is sparse by placing a prior distribution on s that promotes sparsity. A hierarchical prior is imposed on s which has been observed to yield sparse representations for s [9]. First, a zero mean Gaussian prior is defined on s

$$p(s|\boldsymbol{\alpha}) = \prod_{i=1}^{N} \mathcal{N}(s_i|0,\alpha_i^{-1}), \tag{7}$$

with α_i as the inverse of variance of each element. The α_i are further drawn from a Gamma distribution. Next we obtain an expression for the posterior $p(s,\boldsymbol{\alpha},\sigma^2|y)$ by decomposing it as

$$p(s,\boldsymbol{\alpha},\sigma^2|y) = p(s|y,\boldsymbol{\alpha},\sigma^2)p(\boldsymbol{\alpha},\sigma^2|y). \tag{8}$$

The first part of (8) can be computed as

$$p(s|y,\boldsymbol{\alpha},\sigma^2) = \frac{p(y|s,\sigma^2)p(s|\boldsymbol{\alpha})}{p(y|\boldsymbol{\alpha},\sigma^2)}, \tag{9}$$

$$= Cexp(-\frac{1}{2}(s-\mu)^T\Sigma^{-1}(s-\mu)), \tag{10}$$

where the mean and covariance are given by

$$\mu = \sigma^{-2}\Sigma\boldsymbol{\theta}^T y, \tag{11}$$

$$\Sigma = (\sigma^{-2}\boldsymbol{\theta}^T\boldsymbol{\theta} + B)^{-1}, \tag{12}$$

where $B = diag(\alpha_1,\alpha_2,...,\alpha_N)$. The second part of (8) $p(\boldsymbol{\alpha},\sigma^2|y)$ is approximated by a delta distribution at its most probable values [9]. This would mean searching for the maximum values of $p(\boldsymbol{\alpha}, \sigma^2|y)$ and since

$$p(\boldsymbol{\alpha}, \sigma^2|y) \propto p(y|\boldsymbol{\alpha}, \sigma^2)p(\boldsymbol{\alpha})p(\sigma^2), \tag{13}$$

we maximize $p(y|\boldsymbol{\alpha}, \sigma^2)$ as the distributions $p(\boldsymbol{\alpha})$ and $p(\sigma^2)$ are considered uniform because of the choice of parameters for their priors. Now $p(y|\boldsymbol{\alpha}, \sigma^2)$ is computable and given by

$$p(y|\boldsymbol{\alpha}, \sigma^2) = \int p(y|s, \sigma^2)p(s|\boldsymbol{\alpha})ds, \tag{14}$$

$$= C|\sigma^{-2}\boldsymbol{I}+\boldsymbol{\theta}B^{-1}\boldsymbol{\theta}^T|^{-1/2}exp(-y^T(\frac{1}{2}(\sigma^{-2}\boldsymbol{I}+\boldsymbol{\theta}B^{-1}\boldsymbol{\theta}^T)^{-1}y) \tag{15}$$

Now we need to obtain values for $\boldsymbol{\alpha}$ and σ^2 which maximize the function in (15). As described in [9], this can be done by an iterative algorithm. Finally, the posterior distribution for the original signal $\mathbf{x}$ is also a multivariate Gaussian given by

$$E[\mathbf{x}] = \mathcal{B}\mu, \qquad \text{Cov}[\mathbf{x}] = \mathcal{B}\Sigma\mathcal{B}^T. \tag{16}$$

Recent work on the theoretical analysis of Relevance Vector machines has shown that RVM in fact provides a tighter approximation to the l_0 norm based optimization than the l_1 norm optimization technique[11], [12]. The iterative RVM algorithm has also been reported to perform better or at least as good as other l_1 norm based optimization algorithms. Further, the Bayesian framework gives us a handle on the entire distribution rather than a point estimate. This means that the RVM based compressive sensing is more suitable alternative to the more popular l_1 norm minimization based techniques. This also opens the door to investigating how the future measurements can be designed to improve the estimates of the more uncertain points.

IV. Adaptation and Dynamics

The measurement matrix Φ (or $\boldsymbol{\theta}$ in the transform domain) for the compressive sensing techniques described earlier is describes a predetermined and fixed sensing strategy and does not involve adaptation between individual measurements. Since the measurement matrices that we use are not dense but binary (signifying a temporal sequence), we explore whether such matrices can be adaptively designed, i.e. adaptively adding rows to the matrix while ensuring that the estimate obtained keeps getting more accurate. The advantage of such a framework is twofold. First it lets us adaptively design our survey points *during* a static reconstruction of the field. Secondly, it gives us clue on how to incorporate previously learned state history in designing new measurements, *after* the field has changed due to its model dynamics.

A. Adaptive Bayesian Compressive Sensing

The Bayesian CS framework produces a multivariate Gaussian posterior on the signal $\mathbf{x}$ and as measure of the uncertainty in the signal, the differential entropy of $\mathbf{x}$ is given by [7]

$$h_M(\mathbf{x}) = -\int p(\mathbf{x}) \log p(\mathbf{x})d\mathbf{x} \tag{17}$$

$$= -\frac{1}{2}\log|B + \sigma^{-2}\boldsymbol{\theta}^T\boldsymbol{\theta}| + \text{const.} \tag{18}$$

This expression depends on the measurement matrix $\boldsymbol{\theta}$ with M rows, the estimates for $\boldsymbol{\alpha}$ and σ^{-2} and some other constants. Now suppose we wish to add another row r_{M+1} to $\boldsymbol{\theta}$. The effect of adding this new row would be to change the differential entropy of $\mathbf{x}$ to [7]

$$h_{\text{M+1}}(\mathbf{x}) = h_M(\mathbf{x}) - \frac{1}{2}\text{log}(1 + \sigma^{-2}r_{M+1}^T\Sigma r_{M+1}). \tag{19}$$

To minimize the uncertainty in $\mathbf{x}$, we maximizing the term

$$r_{M+1}^T\Sigma r_{M+1} = \text{Var}(y_{M+1}), \tag{20}$$

which shows that the next projection r_{M+1} should be constructed in such a way that the next measurement y_{M+1} comes from the most uncertain subspace. This can be done simply by an eigen-decomposition of Σ and selecting for r_{M+1}^T the eigenvector with the largest eigenvalue [7].

This procedure can be directly adopted in the standard CS framework with dense measurement matrices. However, this creates a problem for constructing sparse binary matrices adaptively because the eigen-analysis of Σ will almost always give *dense* eigenvectors. To take care of this, we look for the nearest sparse binary approximation to the suggested r_{M+1} using standard dot products to the basis frame. We have not done a formal analysis on performance loss due to to this approximation but in practise this works well for most problems we have done simulations on.

B. Incorporating Model Dynamics

The Compressive sensing framework and its Bayesian counterpart are used to estimate an individual sparse signal from its limited measurements. In order to estimate a sequence of signals varying over time, when the dynamics of the changing system acting on the signal are also known, the framework needs to extended to incorporate the model dynamics as well.

As seen in the section on Bayesian Compressive sensing, the framework imposes the sparsity constraint on the signal by assuming a hierarchical prior. The known model dynamics can be viewed as additional information about the signal. The standard way to proceed from here would be to use the *Bayes Filter* algorithm [5] to incorporate this prior. The filter can be implemented using an appropriate estimation technique such as a variant of the Kalman filter or the non-parametric particle filter. For details of this formal development see [8].

In this paper, we take a more practical view point and outline a method that works for our groundwater-like problems. Suppose that the grid model of the field evolves from $\mathbf{x}_t := \psi(x, t)$ to $\mathbf{x}_{t+1} := \psi(x, t+\Delta t)$ following Equation 1. Further, assume that this evolution is a linear transformation of the form

$$\mathbf{x}_{t+1} = A\mathbf{x}_t + q_t, \tag{21}$$

where q_t models the process noise in model dynamics. Let us also assume that initially $\mathbf{x}_t$ has been reconstructed as mean $\hat{\mathbf{x}}$ and covariance Σ_t using the Bayesian compressive framework described above. The standard Bayes's filtering methodology would be to first create a *prediction* $\hat{\mathbf{x}}_{t+1}^p = A\hat{\mathbf{x}}_t$ and then *correct* using the measurement pdf evaluated for the next time step. We instead use the following procedure which follows this predict and correct methodology in spirit but is more explicit in using the adaptation introduced in earlier sections.

We first create pseudo-measurements from the predicted state by

$$y_{t+1}^p = \Phi_t A \hat{\mathbf{x}}_t. \tag{22}$$

Note that Φ_t is the complete binary measurement matrix used for compressive sensing of initial field $\mathbf{x}_t$. Next, we append the actual measurements received (so far) in the current step

$$y_{t+1,M}^a = \Phi_{t+1,M}\mathbf{x}_{t+1}. \tag{23}$$

We then append the two measurement classes together to get

$$y_{t+1,M} = \begin{bmatrix} y_{t+1}^p \\ y_{t+1,M}^a. \end{bmatrix} \tag{24}$$

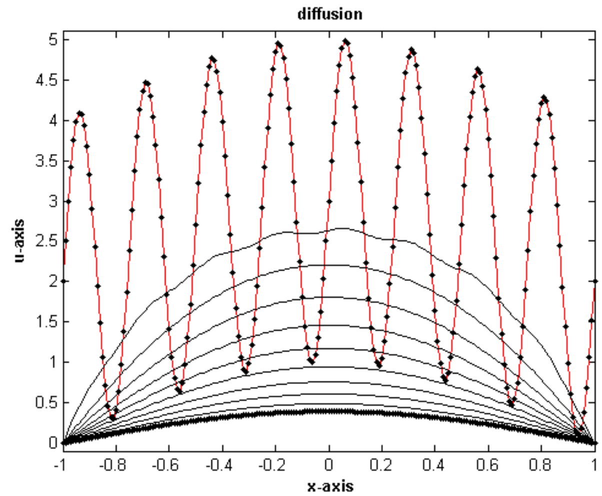
Following the procedure in the previous section, we design the next measurement (or measurement set) by maximizing the variance of $y_{t+1,M}$. This method makes some approximations to the optimal estimation method in that the correction step is not followed to the letter. Practically, we find that this approximation does not prevent us from obtaining reasonably good results. A full investigation of this issue has been taken up in [8].

It should be noted that we currently limit the application of this framework to linear models only. This is because, as described in the section on Bayesian CS, the distribution of the resulting estimate is multivariate Gaussian. To obtain the pseudo measurements, this estimate is propagated to the next time-step and in a linear model this propagated estimate will also be a multivariate Gaussian. However, for the more general non-linear case, we are currently not be able to say much about the distribution of the propagated estimate. However, in principle, the framework should be extendable either through linear approximations or via non-parametric filtering methods.

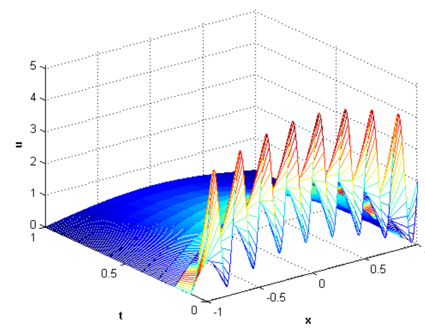
This framework is also expected to give better results with same or less measurements than performing CS with l_1 optimization or the Bayesian CS which are applied on individual signals without assuming any additional knowledge of the system's model based dynamics. This extension of the framework has mainly been done to bring about improvement in results whenever we have this additional information about the system dynamics available.

V. Results and Simulation

We now apply the l_1 norm optimization as well as the Bayesian compressive sensing techniques on two examples where system states are linearly evolving over time and only a limited number of points are available for observation. We first use these techniques to reconstruct individual signals and then use the model incorporating technique and see whether it reduces the number of measurements than those required when prior model information for a signal is not available.



(a)



(b)

Fig. 1. (a) The temperature state of a one dimensional heat rod at each time instant. (b) A 3-D plot showing the transition of the temperature state of the rod. t denotes the time axis and x denotes the spatial axis.

A. One-dimensional Diffusion

The first problem we consider is that of a one dimensional heat equation as a toy model for contaminant diffusion in fluids. Here, the system states are the temperatures of 200 evenly spaced points on a one dimensional rod. Hence, $\mathbf{x}$ is a 200×1 vector of temperature of the rod and s is the vector of DCT transform of x. Our measurement matrix, Φ, is binary because of the problem statement, i.e. the observation points are limited. The temperature at 10 time instants is measured and the above framework is used to estimate s individually for each time instant. Figure 1 shows the original temperature of the rod. As we can see that with time the temperature state of the rod becomes smoother and thus its sparsity pattern in the DCT domain improves.

1) Compressive Sensing using l_1 minimization: We constitute the Binary measurement matrix Φ by randomly selecting around 25 of the 200 points of the rod and consider them as the points that are available for observation. Thus Φ contains the rows from the 200×200 identity matrix corresponding to the locations of the points considered available for measurement. The results by reconstruction using this measurement matrix

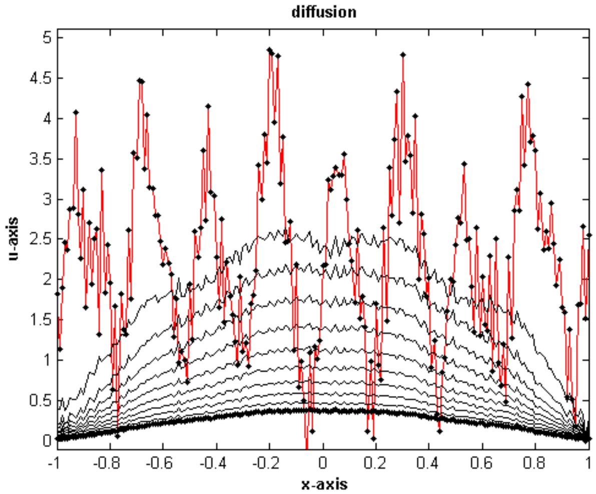


Fig. 2. (a) Estimates of temperature state of the rod using l_1 norm minimization.

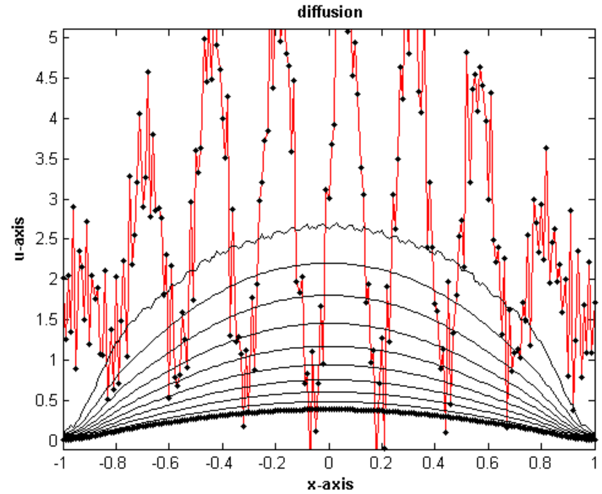


Fig. 3. Estimates of temperature state of the rod using Bayesian compressive sensing.

and the l_1 norm optimization are shown in Figure 2. As can be seen that the initial estimate is not very accurate. This is because as we already mentioned the sparsity pattern of the DCT transform of the temperature states is initially poor but improves over time, and so do the estimates.

2) Bayesian Compressive Sensing: We use the same measurement matrix as in the previous section and the number of points is also kept same. The estimation algorithm used is the faster RVM algorithm. and the MATLAB code for it is available at http://www.ece.duke.edu/ shji/BCS.html. This code defines a function which takes as input the measurement matrix, measurement and the initial noise variance. The output contains the expected values of the non zero elements of the estimates as well as their variances. The initial noise variance is taken to be variance of the measurement divided by 100. The results of this estimation procedure are shown in Figure 3. The results of estimation using Bayesian compressive sensing are slightly better than the l_1 norm based optimization as can be seen in the figure that the estimates of the state temperature are smoother.

3) Using information of Model dynamics: This framework assumes that the model dynamics are known and in the case of the heat rod (21) is linear and is given by a matrix upon discretization of the heat equation for the rod. Using this knowledge, the framework is applied to this example. Since this framework uses the additional prior information about the system's previous state, it should, theoretically, require lesser measurements to give the same results as CS with l_1 optimization and Bayesian CS. The results indicate that the framework indeed requires lesser measurements to give the same performance as the above two methods. The figures shown below were generated by estimates that used only 15 measurements from the one dimensional rod after initially taking 25 measurements to obtain a good estimate of the initial state. Although the measurements were much less than those taken in the previous trials, the final result and error are the same and this shows that the framework does improve on the requirement for the number of measurements by using the prior information of a system's state. The results are shown in Figure 4.

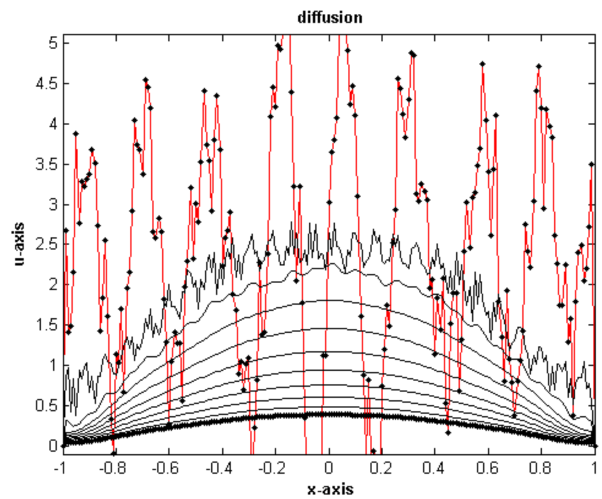


Fig. 4. Estimates of temperature state of the rod using Bayesian compressive sensing along with information about the model dynamics.

Finally we show the relative errors in estimation plotted against time of each of the three techniques in Figure 5. It can be seen that the Bayesian compressive sensing technique and its extension using the model information have smaller errors than compressive sensing based on l_1 norm minimization. The extension to the Bayesian Compressive sensing framework which uses the additional information about how the model is evolving achieves the same estimation accuracy despite measuring the states at lesser points than the Bayesian CS technique (15 compared to 25 points used by BCS).

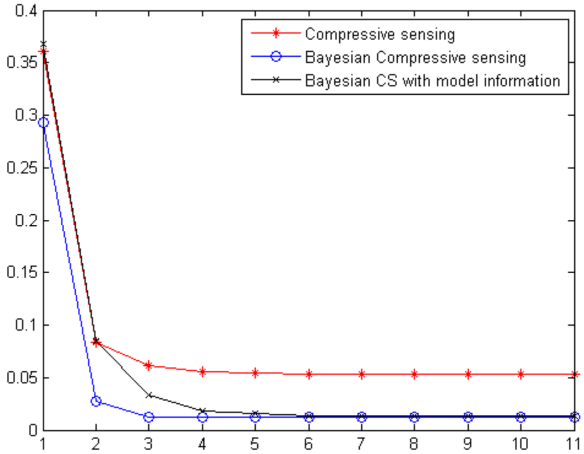


Fig. 5. Relative errors in the estimates of temperature states of the one dimensional rod. Estimates were obtained using l_1 optimization, Bayesian Compressive sensing without model information and Bayesian CS with model information.

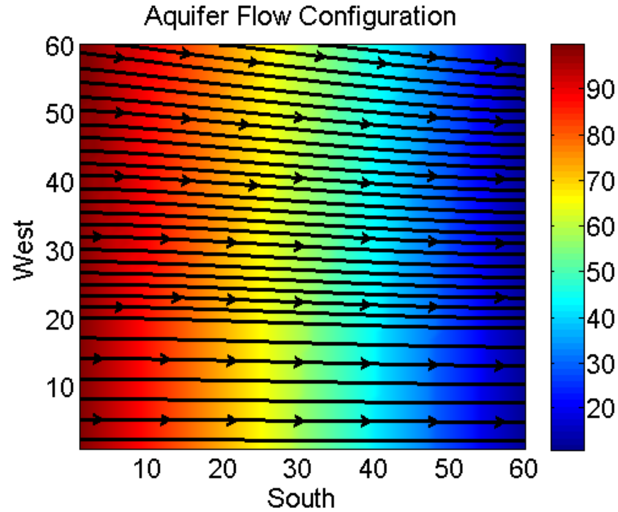


Fig. 6. Groundwater flow configuration in the two-dimensional 60×60 grid. Arrows indicate the direction of flow groundwater. The color indicates the level of groundwater which is highest on the west side and lowest on the eastern end of the region.

B. Groundwater Contaminant in a 2-Dimensional grid

The next problem we consider is that of a groundwater model. We consider a 60×60 two-dimensional grid giving a total of 3600 points or cells. The system state here is the amount of a particular contaminant in the groundwater in each smaller cell. Because such a state would be spatially smooth, the Discrete Cosine Transform is once again a good transform domain for this example as the system state would exhibit a sparse representation in this domain. Assuming system precipitation and permeability in the region, Darcy flow equations can be used to construct a matrix that describes the groundwater flow in the grid. The flow configuration in the grid is shown in Fig 6. We initially assume there is a sharp contaminant plume on the west side of the grid and this plume spreads and moves further in the grid according to the flow configuration over time. The simulation is run for a period of fifty years and the final position of the plume is noted. The initial position of the contaminant plume and snapshots of its movement during this time are shown in 7.

We first perform the Bayesian CS framework on system state without assuming any information about the model dynamics. The system state is measured at about 400 points chosen randomly from the 3600 in the grid. This is based on the assumption that in many real life examples such as this, it is not practically possible to have all the points in a spatial grid available for measurement. This is especially true about groundwater models where each measurement has to be performed by digging up a well or by boring a deep hole in the ground. It would cost too much to perform such a procedure for each point on the grid in question. Thus, about 400 measurements are performed from points chosen randomly and the system state is estimated. The result is shown in Fig 8. Finally we test our proposed extension to the Bayesian CS framework which incorporates the model

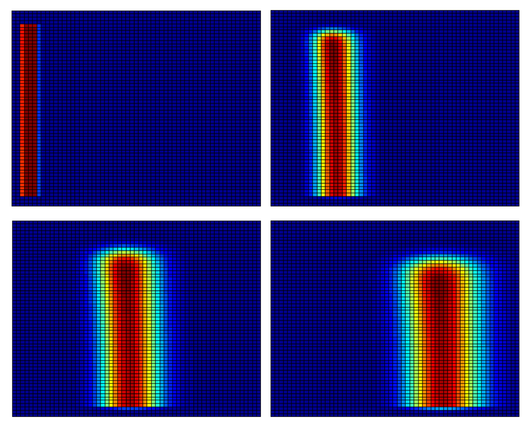


Fig. 7. Simulated 2D plot of the contaminant level over 50 years in the groundwater. The color here indicates the amount of contaminant in a cell in the groundwater.

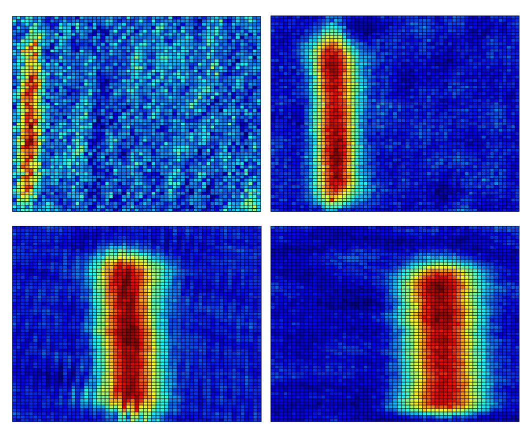


Fig. 8. 2D plot of the estimated contaminant level over 50 years in the groundwater. Estimates are obtained using Bayesian compressive sensing without using knowledge about the aquifer flow configuration

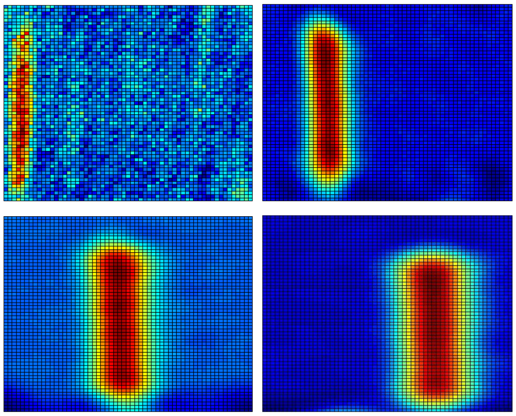


Fig. 9. 2D plot of the estimated contaminant level over 50 years in the groundwater. Estimates are obtained using Bayesian compressive sensing also making use of the additional knowledge about the aquifer flow configuration

dynamics information along with the sparsity constraints.It is important to note the the matrix A_2 in (24) is binary in keeping with the constraint that only a limited number of points are available for measurement. This framework is able to achieve smoother estimates of the system state with around 300 measurements. Thus using the prior information of the system state, this framework has been able to achieve as good an estimate of the contaminant plume after 50 years than the estimate based simply on Bayesian CS. The estimate produced by this framework using only about 300 points from the 3600 in the grid is shown in Fig 9.

VI. Conclusion and Future Directions

In this paper we have discussed techniques for sparse signal recovery and investigated their application on recovery problems where most of the data points are missing from the sample. The techniques we have explored are Compressive sensing based on l_1 norm minimization and a Bayesian Compressive sensing framework based on relevant vector machines. We have seen that the measurement matrices that can be used for the missing data problem are highly sparse binary matrices and their use with both the recovery schemes indicates that the Bayesian framework performs slightly better than the l_1 norm minimization based compressive sensing with the same number of measurements.

We have also developed an extension to the Bayesian Compressive sensing framework in the case of recovering a sequence of system states when the dynamics of the underlying model acting on those states are known to us. This extension utilizes this additional information to achieve the recovery of system state using even lesser measurements than those required by the Bayesian CS framework. The results we have obtained on the two simulated examples where the models are linear indicate that this extension does achieve the same estimation accuracy as the BCS framework but with lesser measurements by utilizing the additional information about the model.

There are more than one avenues related to this work that are worth exploring further. One direction could be a detailed mathematical analysis of the framework that utilizes the model information along with sparse signal recovery. Theoretical results that demonstrate the usefulness of this technique in reducing the number of measurements required for a good enough recovery may be derived. Another area worth exploring is the adaptive techniques for both compressive sensing and its Bayesian counterpart. Although some of the preliminary results we have obtained in this regard indicate that adaptively choosing the next points for measurements generally gives the same estimation errors as randomly choosing the available points beforehand, there is certainly need for this aspect to be explored further. If some adaptive technique does improve the estimates more than random measurements in some cases, it could be used with the extension we have developed to adaptively select the next points for measurements. This would mean that in the practical problem of groundwater survey, for example, we can use the adaptive scheme to direct future surveys towards those points whose measurements, if used, would provide the best reduction in the estimate error.

Acknowledgment

The authors would like to thank Dr. Ibrahim Hoteit and Mr Mohamad El Gharamti at KAUST for providng us with the code to simulate 2D contaminant flow in the groundwater. We also thank Hasan Arshad Nasir who helped us greatly in the initial phase of the project.

References

[1] R. Berinde and P. Indyk, "Sparse recovery using sparse random matrices," *preprint*, 2008.

[2] E. Candes and J. Romberg, "l1-magic: Recovery of sparse signals via convex programming," *URL: www. acm. caltech. edu/l1magic/downloads/l1magic. pdf*, vol. 4, 2005.

[3] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[4] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[5] G. Evensen, *Data assimilation: the ensemble Kalman filter*. Springer, 2009.

[6] B. Jafarpour, V. Goyal, D. McLaughlin, and W. Freeman, "Compressed history matching: Exploiting transform-domain sparsity for regularization of nonlinear dynamic data integration problems," *Mathematical Geosciences*, vol. 42, no. 1, pp. 1–27, 2010.

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

[8] A. Muhammad and Z. Hussain, "Sparse data assimilation in geophysical models," *In Preparation*, 2012.

[9] M. Tipping, "Sparse bayesian learning and the relevance vector machine," *The Journal of Machine Learning Research*, vol. 1, pp. 211–244, 2001.

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

[11] D. Wipf and B. Rao, "l 0-norm minimization for basis selection," *Advances in Neural Information Processing Systems*, vol. 17, 2005.

[12] ——, "Comparing the effects of different weight distributions on finding sparse representations," 2006.

[13] X. Wu and M. Liu, "In-situ soil moisture sensing: measurement scheduling and estimation using compressive sensing," in *Proceedings of the 11th international conference on Information Processing in Sensor Networks*. ACM, 2012, pp. 1–12.