Compression Schemes for Time-Varying Sparse Signals

Sundeep Prabhakar Chepuri and Geert Leus
Faculty of Electrical Engineering, Mathematics and Computer Science,
Delft University of Technology, The Netherlands.
Email: {s.p.chepuri;g.j.t.leus}@tudelft.nl.

***Abstract*—In this paper, we will investigate an adaptive compression scheme for tracking time-varying sparse signals with possibly varying sparsity patterns and/or order. In particular, we will focus on sparse sensing, which enables a completely distributed compression and simplifies the sampling architecture. The sensing matrix is designed at each time step based on the entire history of measurements and known dynamics such that the information gain is maximized. We illustrate the developed theory with a target tracking example. Finally, we provide a few extensions of the proposed framework to include a richer class of sparse signals, e.g., structured sparsity and smoothness.**

***Index Terms*—Structured sensing, sensor selection, sparsity-aware Kalman filter, sparse sensing, adaptive compressed sensing, distributed compression, big data.**

I. Introduction

With the advances in sensor technology, the amount of data produced by pervasive sensors is prohibitively large. A large volume of such data is generated by a variety of applications such as imaging platforms and mobile devices, surveillance cameras and drones, as well as the Internet. The data is often sampled through spatially separated nodes or agents, which then transport the data to a central server for further processing.

It is of crucial importance to start thinking about the inference task we want to perform on the data and to gather only the data that is informative for that specific task. By doing so, we can significantly reduce the sensing cost, data storage, processing overhead, and communications bandwidth. Therefore, the aim is to extend traditional sensing methods to more structured sensing mechanisms for specific signal processing tasks. With such sampling schemes, the sampling rates can be significantly reduced, yet achieving a desired inferential performance.

The problem of choosing the best subset of sensors out of the candidate sensors such that a desired inference performance is achieved is referred to as *sensor selection*. Sensor selection is an experimental design problem, and has been studied in the context of inference tasks like estimation, filtering, and detection [1]–[6] (see references therein). In this paper, we extend the sensor selection framework for non-linear filtering developed in [6] to sampling designs for filtering problems involving structured signals (more generally, filtering problems with equality constraints on the state variables). In particular, we are interested in state sequences that are sparse in nature, which have received a lot of attention in the recent past through *compressive sensing* (CS) [7]. The theory developed under the classical CS framework advocates sensing architectures based on random matrices, which has been proven essential to provide recovery algorithms, reconstruction guarantees, and performance analyses. The CS theory has evolved in recent years to include a much richer class of structured signals. Further, the random sensing architectures are replaced with more structured sensing operators to accommodate practical applications such as sensor networks (e.g., for source localization and field estimation), imaging, and cognitive radio sensing, to list a few. See [8] for a more detailed review on structured CS.

This work was supported in part by STW under the FASTCOM project (10551) and in part by NWO-STW under the VICI program (10382).

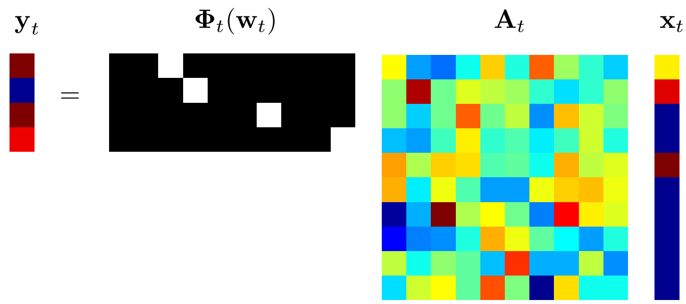


Fig. 1: Illustration of the sparse sensing scheme. Here, the white (black) and colored squares represent a one (zero) and an arbitrary value, respectively.

We consider the problem of *adaptive compressive sensing* of time-varying sparse signals with possibly time-varying sparsity patterns and/or order. This problem has been studied in the past leading to various forms of sparsity-aware filters [9]–[12], and are applied to problems like visual surveillance [13] and target localization [14]. We study the design of sensing matrices for such problems; however, the focus will not be on the signal recovery itself. Sensing matrix design for sparse recovery has been studied in various forms. For example, in [15, Ch. 6], [16] the variance of the distribution from which the (random) sensing matrices are generated is designed such that the average information gain is maximized. The Bayesian CS framework [17] allows to quantify the sparse reconstruction error through the so-called error bars, which again allows to adaptively design the sensing matrices. Both [16] and [17] use experimental design techniques with performance measures like differential entropy to adaptively learn the sensing matrix starting from a random matrix. In [18], a greedy algorithm based on a submodular performance measure has been proposed for sensing operator design for a signal lying in the union of subspaces. However, the sensing design schemes discussed above are mostly limited to time-invariant signals and/or systems without any state-space representation. Hence, they are not adaptive in the true sense, and are not meant for tracking the signal variation over space and/or time.

In this paper, the sensing matrix is designed at each time step based on the entire history of measurements and known dynamics described through a state-space model. The sensing matrices are parameterized through a sparse Boolean vector. Hence, the resulting sampling architecture is referred to as *sparse sensing* (see the illustration in Fig. 1). The sensing matrix is deterministic, and hence it is more favorable for practical scenarios including distributed sampling that is crucial for sensor network applications. The non-zero entries of the Boolean vector are determined by optimizing the *a posteriori* error. Thus, it maximizes the information gain or equivalently reduces uncertainty. The proposed offline design problem is convex in nature and can be readily solved using off-the-shelf software. We also discuss a few extensions of the proposed framework to include general structured signals, such as group sparsity and smoothness.

The notation used in this paper can be described as follows. Upper (lower) bold face letters are used for matrices (column vectors). $(\cdot)^T$ denotes transposition. $\mathrm{diag}(\cdot)$ refers to a block diagonal matrix with the elements in its argument on the main diagonal. $\mathbf{1}_N$ ($\mathbf{0}_N$) denotes the $N \times 1$ vector of ones (zeros). $\mathbf{I}_N$ is an identity matrix of size N. $\mathbb{E}\{\cdot\}$ denotes the expectation operation. $\mathrm{tr}\{\cdot\}$ is the matrix trace operator. $\det\{\cdot\}$ is the matrix determinant. $\lambda_{\min}\{\mathbf{A}\}$ ($\lambda_{\max}\{\mathbf{A}\}$) denotes the minimum (maximum) eigenvalue of a symmetric matrix $\mathbf{A}$.

II. Problem Statement

Assume that the time-varying vector of interest $\mathbf{z}_t \in \mathbb{R}^M$ is compressible in some known linear basis denoted by an $M \times M$ matrix $\mathbf{A}_t$. In other words, we can express $\mathbf{z}_t = \mathbf{A}_t\mathbf{x}_t$, where $\mathbf{x}_t$ has just a few non-zero coefficients, i.e., $\|\mathbf{x}_t\|_0 \ll M$ ($\|\cdot\|_0$ counts the non-zero entries of its argument). Under the assumption that the parameter vector $\mathbf{x}_t$ is sparse, CS theory asserts an exact recovery of $\mathbf{x}_t$ from observations which are typically much smaller than M, i.e, signals acquired via a linear compression matrix. In this paper, we are interested in designing a time-varying compression matrix as well as determining the optimal compression rate to reach a desired information gain or mean-squared error.

The unknown parameter $\mathbf{x}_t$ follows a linear model corrupted by additive noise:

$$\mathbf{y}_t = \mathbf{\Phi}_t\mathbf{z}_t + \mathbf{n}_t = \mathbf{\Phi}_t\mathbf{A}_t\mathbf{x}_t + \mathbf{n}_t, \tag{1}$$

where the (spatial and/or temporal) measurements at a temporal block t are stacked in the measurement vector $\mathbf{y}_t = [y_{t,1}, y_{t,2}, \ldots, y_{t,N}]^T \in \mathbb{R}^N$, and $\mathbf{\Phi}_t \in \mathbb{R}^{N\times M}$ denotes the sampling matrix. For $N \ll M$, the sampling matrix will be a compression matrix, and the measurement vector will be much shorter than $\mathbf{x}_t$. The additive noise vector $\mathbf{n}_t = [n_{t,1}, n_{t,2}, \ldots, n_{t,N}]^T \in \mathbb{R}^N$ is assumed to be zero-mean with covariance matrix $\mathbf{R}_{\mathbf{n}_t} = \sigma^2\mathbf{I}_N$. Note that the sampling matrix $\mathbf{\Phi}_t$ and sparsity pattern (including the support size) of the vector $\mathbf{x}_t$ can both be time-varying.

The unknown sparse parameter is assumed to obey the following dynamical model

$$\mathbf{x}_t = h_t(\mathbf{x}_{t-1}, \mathbf{v}_{t-1}), \tag{2}$$

where the process noise is denoted by $\mathbf{v}_{t-1} \in \mathbb{R}^{M\times 1}$, which accounts for any unmodeled dynamics. The evolution of the sparse vectors $\{\mathbf{x}_t, t \in \mathbb{N}\}$ is governed by a non-linear function $h_t : \mathbb{R}^M \times \mathbb{R}^M \to \mathbb{R}^M$. Here, we model $\mathbf{v}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_{\mathbf{v}_t})$, where $\mathbf{R}_{\mathbf{v}_t} \in \mathbb{R}^{M\times M}$ represents the covariance matrix of $\mathbf{v}_t$. Alternatively, the evolution of the time-varying sparse sequence (2) can be described using a pseudo-measurement formulation [9], [14], [19]. More specifically, it is assumed that $\mathbf{x}_t$ evolves according to the following model

$$\text{dynamics:} \quad \mathbf{x}_t = \mathbf{H}_t\mathbf{x}_{t-1} + \mathbf{u}_t; \tag{3a}$$

$$\text{pseudo-measurement:} \quad 0 = g(\mathbf{x}_t) + e_t, \tag{3b}$$

where $\mathbf{H}_t$ is an $M \times M$ state-transition matrix, $\mathbf{u}_t \in \mathbb{R}^M$ is the process noise, $g(\mathbf{x}_t)$ is a sparsity-controlling convex function, and e_t is zero-mean unit-variance noise. For example, $g(\cdot)$ can include any one of the well-known approximations of the ℓ_0(-quasi) norm such as the ℓ_1-norm $\sum_{m=1}^M |x_{m,t}|$, logarithmic function $\sum_{m=1}^M \log(|x_{m,t}| + \delta)$ with $\delta > 0$, or the inverse Gaussian function $\sum_{m=1}^M \left(1 - \exp\left(-\frac{x_{m,t}^2}{2\sigma_g^2}\right)\right)$ with tuning parameter σ_g^2. Here, $x_{m,t}$ denotes the mth entry of $\mathbf{x}_t$. Henceforth, we will restrict ourselves to (3) instead of (2) because of the generalization it offers to accommodate a much richer class of structured signals as discussed later on in Section VI.

Remark 1 (Static case)**.** *A specialization of* (3) *is the static case, i.e.,* $\{\mathbf{x}_t, t \in \mathbb{N}\}$ *is time-invariant. This is obtained with* $\mathbf{H}_t = \mathbf{I}_M$ *and* $\mathbf{R}_{\mathbf{u}_t} = \mathbf{0}_{M\times M}$ *for* $t = 1, 2, \ldots, T$.

We consider the following *adaptive sparse sensing* problem. For each $t = 1, 2, \ldots, T$, design a deterministic sparse compression matrix $\mathbf{\Phi}_t$ based on the entire history of measurements up to that point $\{\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{t-1}\}$ using the state-space model (1) and (3). The matrices $\{\mathbf{\Phi}_t, t \in \mathbb{N}\}$ are chosen from a predefined set, i.e., $\mathbf{\Phi}_t$ is parameterized by a sensing vector $\mathbf{w}_t \in \{0, 1\}^M$ that has to be designed. More specifically, the sensing matrix takes the following form

$$\mathbf{\Phi}_t = \mathrm{diag}_{\mathrm{r}}(\mathbf{w}_t),$$

where the notation $\mathrm{diag}_{\mathrm{r}}(\cdot)$ represents a diagonal matrix with the argument on its diagonal but with the all-zero rows removed. An illustration of the sparse sensing scheme is shown in Fig. 1. In this work, we are interested in designing a sequence of sparse vectors $\{\mathbf{w}_t, t \in \mathbb{N}\}$ (and, hence a sequence of matrices $\{\mathbf{\Phi}_t(\mathbf{w}_t), t \in \mathbb{N}\}$) that results in a desired information gain or mean-squared-error. Note that we are basically replacing the random measurement operation traditionally used in the CS framework with a deterministic and sparse sensing operation, which is more favorable for practical implementation.

III. Optimization Criterion

Suppose that the estimate $\widehat{\mathbf{x}}_{t-1|t-1}$ and its covariance matrix $\mathbf{P}_{t-1|t-1}$ are available from the previous time step. At current time step t, the prediction and its covariance are given as

$$\begin{aligned}\widehat{\mathbf{x}}_{t|t-1} &= \mathbf{H}_t \widehat{\mathbf{x}}_{t-1|t-1} \\ \mathbf{P}_{t|t-1} &= \mathbf{H}_t \mathbf{P}_{t-1|t-1} \mathbf{H}_t^T + \mathbf{R}_{\mathbf{u}_t}.\end{aligned} \tag{4}$$

Given the state-space equations (1) and (3a), without any sparsity constraint, the update step of the standard Kalman filter can be written as the following weighted least-squares problem

$$\begin{aligned}\widehat{\mathbf{x}}_{t|t} = \arg\min_{\mathbf{x}_t} \quad & \|\widehat{\mathbf{x}}_{t|t-1} - \mathbf{x}_t\|^2_{\mathbf{P}^{-1}_{t|t-1}} \\ & + \|\mathbf{y}_t - \mathbf{\Phi}_t(\mathbf{w}_t)\mathbf{A}_t\mathbf{x}_t\|^2_{\mathbf{R}^{-1}_{\mathbf{n}_t}},\end{aligned} \tag{5}$$

where $\|\mathbf{x}\|^2_{\mathbf{A}} = \mathbf{x}^T\mathbf{A}\mathbf{x}$. The *a posteriori* error covariance can be written recursively for $t = 1, 2, \ldots, T$ as

$$\mathbf{P}_{t|t} := \left(\mathbf{P}^{-1}_{t|t-1} + \frac{1}{\sigma^2}\sum_{m=1}^{M} w_{m,t}\mathbf{a}_{m,t}\mathbf{a}^T_{m,t}\right)^{-1}, \tag{6}$$

where $\mathbf{P}_{t|t-1} = \mathbf{R}_{\mathbf{u}_t} + \mathbf{H}_t\mathbf{P}^{-1}_{t-1|t-1}\mathbf{H}^T_t$ with $\mathbf{R}_{\mathbf{u}_t}$ being the covariance matrix of $\mathbf{u}_t$, $\{\mathbf{a}_{m,t} \in \mathbb{R}^{M\times 1}, m = 1, 2, \ldots, M\}$ are the rows of $\mathbf{A}_t$, and $\mathbf{w}_t = [w_{1,t}, w_{2,t}, \ldots, w_{M,t}]^T$ is the sampling operator.

The linear system (1) and (3a) is observable for $\mathbf{\Phi}_t = \mathbf{I}_N$, and it can be solved using the celebrated Kalman filter (KF) algorithm (5) and (6). Due to the compression, the number of unobservable modes increases and the conventional KF in no more useful, unless the inherent sparsity of the state sequence is taken into account. We do this through an (independent) extra pseudo-measurement (3b), which then modifies the update equation (5) to the following non-linear least-squares problem

$$\begin{aligned}\widehat{\mathbf{x}}_{t|t} = \arg\min_{\mathbf{x}_t} \quad & \|\widehat{\mathbf{x}}_{t|t-1} - \mathbf{x}_t\|^2_{\mathbf{P}^{-1}_{t|t-1}} \\ & + \|\mathbf{y}_t - \mathbf{\Phi}_t(\mathbf{w}_t)\mathbf{A}_t\mathbf{x}_t\|^2_{\mathbf{R}^{-1}_{\mathbf{n}_t}} + g^2(\mathbf{x}_t),\end{aligned} \tag{7}$$

The above optimization problem can be solved using a Gauss-Newton algorithm leading to an iterative extended Kalman filter implementation; see [9], [14] for more details. Furthermore, the *a posteriori* error covariance (6) is modified to

$$\begin{aligned}\mathbf{P}_{t|t} := \Bigg(&\mathbf{P}^{-1}_{t|t-1} + \partial g(\widehat{\mathbf{x}}_{t|t-1})\partial g(\widehat{\mathbf{x}}_{t|t-1})^T \\ &+ \frac{1}{\sigma^2}\sum_{m=1}^{M} w_{m,t}\mathbf{a}_{m,t}\mathbf{a}^T_{m,t}\Bigg)^{-1},\end{aligned} \tag{8}$$

where $\partial g(\widehat{\mathbf{x}}_{t|t-1}) \in \mathbb{R}^M$ is the (sub)gradient of $g(\mathbf{x}_t)$ towards $\mathbf{x}_t$ evaluated at $\widehat{\mathbf{x}}_{t|t-1}$. The second term in the above expression is due to the virtual measurement (3b) that accounts for the sparsity of the state sequences $\{\mathbf{x}_t, t \in \mathbb{N}\}$. For example, using $g(\mathbf{x}_t) := \sqrt{2\gamma_t\|\mathbf{x}_t\|_1}$ with tuning parameter γ_t in (8), we obtain the following performance measure

$$\begin{aligned}\mathbf{P}_{t|t}(\mathbf{w}_t) := \Bigg(&\mathbf{P}^{-1}_{t|t-1} + \frac{\gamma_t}{2\|\widehat{\mathbf{x}}_{t|t-1}\|_1}\bar{\mathbf{x}}_{t|t-1}\bar{\mathbf{x}}^T_{t|t-1} \\ &+ \frac{1}{\sigma^2}\sum_{m=1}^{M} w_{m,t}\mathbf{a}_{m,t}\mathbf{a}^T_{m,t}\Bigg)^{-1},\end{aligned} \tag{9}$$

where $\bar{\mathbf{x}}_{t|t-1} = \left[\frac{\widehat{x}_{1,t|t-1}}{|\widehat{x}_{1,t|t-1}|}, \frac{\widehat{x}_{2,t|t-1}}{|\widehat{x}_{2,t|t-1}|}, \cdots, \frac{\widehat{x}_{M,t|t-1}}{|\widehat{x}_{M,t|t-1}|}\right]^T$ with $\widehat{\mathbf{x}}_{t|t-1} = [\widehat{x}_{1,t|t-1}, \ldots, \widehat{x}_{M,t|t-1}]^T$. The focus however is not on solving (7) itself, but on designing $\mathbf{w}_t$ to acquire $\mathbf{y}_t$ based on the information available one step ahead in time. We next propose solvers for designing the sparse sensing matrices based on the optimization criterion (9) and for a specific desired performance.

IV. Sparse Sensing Design

The sensing matrix design is determined by evaluating scalar functions of the *a posteriori* error covariance (9). Some of the prominent scalar functions are related to A-optimality (signifies the mean-squared-error), E-optimality (signifies the worst case error), and D-optimality (signifies the entropy), and are respectively given as $f_t(\mathbf{w}_t) := \mathrm{tr}\{\mathbf{P}_{t|t}\}$, $f_t(\mathbf{w}_t) := \lambda_{\min}\{\mathbf{P}_{t|t}\}$, and $f_t(\mathbf{w}_t) := \det\{\mathbf{P}_{t|t}\}$. Here, the performance measure $f_t(\cdot)$ is time-varying. All the above criteria quantify reasonably the "largeness" of the information content (or reduction in uncertainty). Hence, the sensing operation is designed such that one of these functions as well as the cardinality of $\mathbf{w}_t$ are jointly minimized, where the cardinality of $\mathbf{w}_t$ represents the compression rate. In other words, the compression rate increases with the number of non-zero entries of $\mathbf{w}_t$.

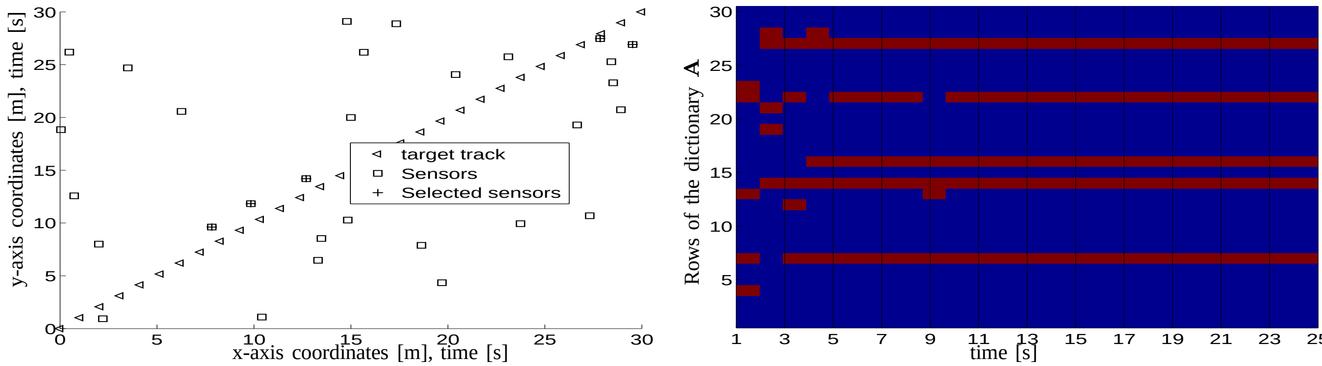
Mathematically, in adaptive sparse sensing, at each time step $t = 1, 2, \ldots, T$, we solve the following optimization problem

$$\begin{aligned}\arg\min_{\mathbf{w}_t} \quad & \|\mathbf{w}_t\|_0 \\ \text{s.to} \quad & f_t(\mathbf{w}_t) \leq \lambda, \\ & \mathbf{w}_t \in \{0, 1\}^M,\end{aligned} \tag{10}$$

where the threshold λ specifies the desired accuracy and also controls the compression rate. The optimization problem in (10) is a non-convex Boolean problem. We simplify (10) by replacing the ℓ_0(-quasi) norm with its best convex approximation, i.e., the ℓ_1-norm given as $\mathbf{1}^T_M\mathbf{w}_t$ and the Boolean $\{0, 1\}$ constraint with a box constraint $[0, 1]$. However, the performance measure $f_t(\mathbf{w}_t)$ is convex in $\mathbf{w}_t$ [1]. Thus, the relaxed optimization problem is given as

$$\begin{aligned}\arg\min_{\mathbf{w}_t} \quad & \mathbf{1}^T_M\mathbf{w}_t \\ \text{s.to} \quad & f_t(\mathbf{w}_t) \leq \lambda, \\ & 0 \leq w_{m,t} \leq 1, m = 1, 2, \ldots, M.\end{aligned} \tag{11}$$

The above optimization problem is convex in $\mathbf{w}_t$. However, the solution of (11) is not yet Boolean, and the approximate Boolean solution has to be recovered. This can be done either by deterministic or randomized rounding [4]. The relaxed



(a) (b)

Fig. 2: Tracking a target using a grid-based model with $M = 30$ and $N = 5$. (a) A target is moving along the straight line $\theta_{x,t} = \theta_{y,t} = t$, i.e., it moves with a constant velocity of 1 m/s. Selected sensors shown correspond to $t = 25$ s. (b) The solution path illustrating the selected rows of the dictionary $\mathbf{A}$ for $t = 1, 2, \ldots, 25$ s.

optimization problem can be solved using readily available convex optimization solvers like CVX [20] or SeDuMi [21]. We underline that the formulation (11) will also optimize the number of rows of $\mathbf{\Phi}_t$. In case a specific compression rate is desired (i.e., N is known *a priori*), we solve the following equivalent problem:

$$\begin{aligned} &\arg\min_{\mathbf{w}_t} \; f_t(\mathbf{w}_t) \\ &\text{s.to} \quad \mathbf{1}_M^T \mathbf{w}_t = N, \\ &\qquad\quad 0 \le w_{m,t} \le 1, m = 1, 2, \ldots, M. \end{aligned} \tag{12}$$

Note that other approximations for $\|\mathbf{w}_t\|_0$ can also be considered, such as the sum-of-logarithms $\sum_{m=1}^{M} \log(w_{m,t} + \delta)$ with a small $\delta > 0$ [4].

V. Example: Target Tracking

In this section, we illustrate the developed theory with the following target tracking example. Let $\boldsymbol{\theta}_t = [\theta_{x,t}, \theta_{y,t}]^T \in \mathbb{R}^2$ denote the position of the target at time instance t and $\mathbf{p}_m \in \mathbb{R}^2$ denote the position of the mth virtual sensor. Let us assume that there are M locations where we can place these sensors. The candidate sensors are capable of measuring the signal strength according to the following model $z_{m,t} = a_{m,t}(\boldsymbol{\theta}_t)$ for $m = 1, 2, \ldots, M$, where

$$a_{m,t}(\boldsymbol{\theta}_t) = \frac{\beta s}{\beta + \|\boldsymbol{\theta}_t - \mathbf{p}_m\|_2^2} \tag{13}$$

with a constant $\beta > 0$. Here, s denotes the signal strength. We linearize (13) around M grid points $\{\mathbf{g}_m\}_{m=1}^{M}$, where the target could be potentially located. In this case, we are interested in tracking a target moving along a straight line $\theta_{x,t} = \theta_{y,t} = t$ as shown in Fig. 2a. As a result, we arrive at the linear grid-based model given by

$$z_{m,t} = \mathbf{a}_m^T \mathbf{x}_t, \quad m = 1, 2, \ldots, M,$$

where $\mathbf{a}_m = [a_{m,t}(\mathbf{g}_1), a_{m,t}(\mathbf{g}_2), \ldots, a_{m,t}(\mathbf{g}_M)]^T \in \mathbb{R}^M$ is time-invariant, but $\mathbf{x}_t$ is time-varying. All the entries of the vector $\mathbf{x}_t \in \mathbb{R}^M$ are equal to zero except for the mth entry, $x_{m,t}$, which is equal to the target signal strength s at time t if and only if the target is located at the mth grid point, i.e., $\boldsymbol{\theta}_t = \mathbf{g}_m$. By letting $\mathbf{z}_t = [z_{1,t}, z_{2,t}, \ldots, z_{M,t}]^T$ and the dictionary $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_M]^T \in \mathbb{R}^{M \times M}$, we have $\mathbf{z}_t = \mathbf{A}\mathbf{x}_t$. Note that the number of grid points can be much larger than the number of sensors.

The sparsity of $\mathbf{x}_t$ allows us to uniquely recover $\mathbf{z}_t$ from its compressed version. This is the central idea behind the CS theory. Recalling the measurement model (1), i.e.,

$$\mathbf{y}_t = \mathbf{\Phi}_t(\mathbf{w}_t)\mathbf{z}_t + \mathbf{n}_t,$$

we re-emphasize the advantage of *sparse sensing* (cf. Fig. 1) as follows. In scenarios like the one considered here, compression via random linear projections would still need all the M sensors with no reduction in the sensing and communications cost. On the contrary, sparse sensing enables a completely decentralized sensing, and it needs only $N \ll M$ sensors. Moreover, such spatial sampling schemes lead to (CS-based) sensor placement pertinent to applications like medical imaging, visual surveillance, radar, cognitive radio, to list a few.

The target follows a constant velocity model defined by

$$\mathbf{x}_t = \mathbf{H}\mathbf{x}_{t-1} + \mathbf{u}_t,$$

where the entries of the initial vector $\mathbf{x}_0$ are all zero except for the first entry $x_{1,1} = s$. Here, the state-transition matrix is a shift matrix, i.e.,

$$\mathbf{H} = \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 1 & 0 \end{bmatrix} \in \mathbb{R}^{M \times M}.$$

In this example, the sparsity pattern is time-varying, but the sparsity order is fixed. We stress here that the proposed framework is not limited to signals with a fixed sparsity order. We use the following parameters in the simulations: The number of grid points/candidate sensors $M = 30$, $N = 5$, $\beta = 100$, and $s = 10$. The sensors are (virtually) deployed uniformly at random within a 30×30 m^2 surveillance area as shown in Fig. 2a. We choose potential target grid points $\mathbf{g}_m = [m, m]^T$ for $m = 1, 2, \ldots, 30$. Furthermore, we use

$f_t(\mathbf{w}_t) := \mathrm{tr}\{\mathbf{P}_{t|t}^{-1}\}$, $\widehat{\mathbf{x}}_{0|0} = \mathbf{1}_M$, $\mathbf{P}_{0|0} = \mathbf{I}_M$, $\mathbf{R}_{\mathbf{u}_t} = 0.01\mathbf{I}_M$, and $\sigma = 10^{-3}$. We compute γ_t using the method described in [14] and λ is chosen such that $N = 5$ sensors are selected at each time step.

The relaxed optimization problem is solved using CVX, which internally calls SeDuMi. The update step of the sparsity-aware Kalman filter is computed using (7), while we use (9) for the covariance update and (4) for prediction. Recall that we design the sensing matrix $\mathbf{\Phi}_t(\mathbf{w}_t)$ one step ahead by optimizing a scalar function of (9). Fig. 2b illustrates the solution path obtained by solving (11) for $t = 1, 2, \ldots, 25$ s. The Boolean solution is recovered using deterministic rounding. The sensors selected for time step $t = 25$ s is also shown in Fig. 2a. In this example, the same subset of sensors are selected for $t > 10$ because the matrices $\mathbf{A}$ and $\mathbf{H}$, and the sparsity order are not changing with time.

VI. Extensions

In this section, we highlight some important generalizations of the proposed framework, which are often studied together with the CS framework. The sparsity prior can be extended to a much broader class of structured signals, including structured sparse signals (or block-sparse signals) [22], smoothness (i.e., sparsity of the coefficients and also sparsity of their differences) [23], to list a few. Depending on the structure of the state, the $g(\mathbf{x}_t)$ has to be modified accordingly. More specifically, for structured sparse signals we use a regularizer that accounts for block sparsity, i.e.,

$$g(\mathbf{x}_t) := 2\gamma_t \sum_{i=1}^{G} \|\mathbf{x}_i\|_2,$$

where the state vector $\mathbf{x}_t$ is grouped into G subvectors each of length N/G as $\mathbf{x}_t = [\mathbf{x}_{1,t}^T, \mathbf{x}_{2,t}^T, \ldots, \mathbf{x}_{G,t}^T]^T$. Similarly, for signal smoothness, we use the regularizer

$$g(\mathbf{x}_t) := \gamma_{1,t}\|\mathbf{x}_t\|_1 + \gamma_{2,t} \sum_{m=1}^{M-1} |x_{m,t} - x_{m-1,t}|,$$

where $\gamma_{1,t}$ and $\gamma_{2,t}$ are tuning parameters. Note that using the above regularizers in (7), equations (7) and (8) together will form the update step of a general *structured-sparsity-aware Kalman filter*.

VII. Conclusions

In the era of big data, it is very crucial to design sensing operators keeping in mind the specific task we want to perform on the acquired data. Thus, only the informative data has to be acquired such that the inferential performance is still acceptable. The sensing architectures for dimensionality reduction, especially under the classical CS framework are mostly based on random matrices. In this paper, we have developed a framework for sensing matrix design for time-varying sparse signals. In particular, we have considered the design of a deterministic and sparse sensing matrix, which is essential for decentralized compression. The proposed solvers are convex in nature and can be solved using off-the-shelf software. We have also provided some extensions of the proposed framework to include a much richer class of sparse signals (e.g., block sparse, smoothness), which also leads to a structured-sparsity-aware Kalman filter.

References

[1] S. P. Chepuri and G. Leus, "Sensor selection for estimation, filtering, and detection," in *Proc. of International Conference on Signal Processing and Communications (SPCOM), 2014*, July 2014, pp. 1–5.

[2] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in gaussian processes: Theory, efficient algorithms and empirical studies," *The Journal of Machine Learning Research*, vol. 9, pp. 235–284, Feb. 2008.

[3] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, Feb. 2009.

[4] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Trans. Signal Process.*, Dec. 2014.

[5] D. Bajovic, B. Sinopoli, and J. Xavier, "Sensor selection for event detection in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 59, no. 10, pp. 4938–4953, Oct 2011.

[6] S. P. Chepuri and G. Leus, "Sparsity-promoting adaptive sensor selection for non-linear filtering," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2014, pp. 5100–5104.

[7] R. G. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, 2007.

[8] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4053–4085, 2011.

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

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

[11] D. Angelosante, G. B. Giannakis, and E. Grossi, "Compressed sensing of time-varying signals," in *Proc. of 16th International Conference on Digital Signal Processing, 2009*. IEEE, 2009, pp. 1–8.

[12] D. M. Malioutov, S. R. Sanghavi, and A. S. Willsky, "Sequential compressed sensing," *IEEE J. Sel. Topics in Signal Process.*, vol. 4, no. 2, pp. 435–444, 2010.

[13] G. Warnell, D. Reddy, and R. Chellappa, "Adaptive rate compressive sensing for background subtraction," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2012*. IEEE, 2012, pp. 1477–1480.

[14] S. Farahmand, G. B. Giannakis, G. Leus, and Z. Tian, "Tracking target signal strengths on a grid using sparsity," *EURASIP Journal on Advances in Signal Processing*, vol. 2014, no. 1, p. 7, 2014.

[15] Y. C. Eldar and G. Kutyniok, *Compressed sensing: theory and applications*. Cambridge University Press, 2012.

[16] J. Haupt, R. Nowak, and R. Castro, "Adaptive sensing for sparse signal recovery," in *Proc. of IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop, 2009. DSP/SPE 2009.* IEEE, 2009, pp. 702–707.

[17] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, 2008.

[18] D. El Badawy, J. Ranieri, and M. Vetterli, "Near-optimal sensor placement for signals lying in a union of subspaces," in *Proc. of 22nd European Signal Processing Conference (EUSIPCO 2014)*, 2014.

[19] S. Julier and J. LaViola, "On kalman filtering with nonlinear equality constraints," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2774–2784, June 2007.

[20] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http://cvxr.com/cvx, Sep. 2012.

[21] J. F. Sturm, "Using SeDuMi 1.02, a matlab toolbox for optimization over symmetric cones," *Optimization methods and software*, vol. 11, no. 1-4, pp. 625–653, 1999.

[22] J. Friedman, T. Hastie, and R. Tibshirani, "A note on the group lasso and a sparse group lasso," *arXiv preprint arXiv:1001.0736*, 2010.

[23] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and smoothness via the fused lasso," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 67, no. 1, pp. 91–108, 2005.