Sparse Sensing for Estimation with Correlated Observations

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

***Abstract*—We focus on discrete sparse sensing for non-linear parameter estimation with colored Gaussian observations. In particular, we design offline sparse samplers to reduce the sensing cost as well as to reduce the storage and communications requirements, yet achieving a desired estimation accuracy. We optimize scalar functions of the Cramér-Rao bound matrix, which we use as the inference performance metric to design the sparse samplers of interest via a convex program. The sampler design does not require the actual measurements, however it needs the model parameters to be perfectly known. The proposed approach is illustrated with a sensor placement example.**

***Index Terms*—Sparse sensing, sensor selection, sensor placement, dependent observations, non-linear least squares.**

I. Introduction

Sensor networks are widely used in a variety of applications such as infrastructure (e.g., buildings, railway tracks, bridges) monitoring, healthcare, environmental monitoring, and safety and security, to name a few. Typically, sensors are deployed spatially and the data gathered by them are transported to a central location for data processing, i.e., for inference. If the inference task is known beforehand, then the cost of sensing as well as the resulting data storage and communications costs can be seriously reduced. The sensing cost includes the hardware cost due to the number of sensors and the physical space they occupy when installed.

In this work, the focus is on the design of sparse sensing structures to perform sensor selection or placement for non-linear parameter estimation. Sensor selection (placement) is the problem of choosing the best subset of sensors (locations) out of a large pool of candidate sensors (locations) in order to reach a desired estimation accuracy. In other words, we are interested in designing optimal sparse spatial or temporal samplers to reduce the data to be acquired. For example, for indoor localization systems this means the number of access points as well as their sampling rates can be significantly reduced for a given target detection probability or positioning accuracy.

In recent years, sensor selection and management has received a significant amount of attention for various signal processing problems such as control and estimation [1]–[10] and detection [11]–[13]. In [1] and [2], the sensor selection problem for linear parameter estimation with uncorrelated Gaussian noise was solved via greedy submodular maximization and convex optimization, respectively. In [3], this problem was generalized to nonlinear measurement models with arbitrary yet conditionally independent data distributions, where convex optimization methods were employed to determine the best subset of sensors such that a desired inference performance is guaranteed (e.g., specified by the Cramér-Rao bound).

The assumption in [2], [3] is that the observations are conditionally independent. This is reasonable if the sensors are solely responsible for the noise in the observations, for example, due to the internal thermal noise. If the observation signal itself is stochastic in nature or if the observations are subject to external noises or interference, then the independence assumption will be too idealistic. As a consequence, the additive property (this result was central to the solvers developed in [3]) of the Fisher information matrix (FIM) under sparse sensing is no more preserved. This makes the sensor selection problem with correlated observations even more challenging.

Existing works, e.g., [14], [15] focus on linear models in colored Gaussian noise, but with an approximate performance metric. That is, they rely on an approximate expression for the mean squared error, where the noise from the sensors that are not selected is spread across the selected sensors. However, the noise from the sensors that are not selected should not have any impact.

On the other hand, we consider an exact expression for the inference performance metric in this work. More specifically, we extend the framework of [3] to solve the sensor selection problem for nonlinear estimation problems with colored Gaussian observations. Similar to [3], we will use the Cramér-Rao bound as an inference performance metric to design sparse samplers. We propose a decomposition of the noise covariance matrix that allows us to express the inference performance metric as a convex function of the sampler. The sparse samplers can be obtained by solving a semidefinite program. The proposed approach is valid for arbitrary covariance matrix (i.e., for high or low correlation values), and uses only model information (and not actual measurements).

Notation: Upper (lower) bold face letters are used for matrices (column vectors). $(\cdot)^T$ denotes transposition. $\mathbf{1}_N$ ($\mathbf{0}_N$) 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 $\mathbf{A}$. $\mathbf{A}\succeq\mathbf{B}$ means that $\mathbf{A}-\mathbf{B}$ is a positive semidefinite matrix. $\|\boldsymbol{w}\|_0$ denotes the ℓ_0 norm of

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

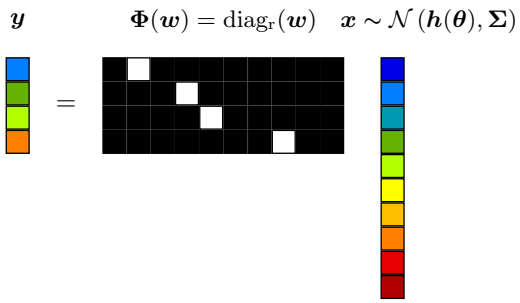

Fig. 1: Discrete sparse sensing scheme. Here, a white (black) and colored square represents a one (zero) and an arbitrary value, respectively.

$\boldsymbol{w}$. $\mathbb{S}^N$ ($\mathbb{S}^N_+$) denotes the set of symmetric (symmetric positive semi-definite) matrices of size $N \times N$.

II. Sensing model

Suppose the unknown vector $\boldsymbol{\theta} \in \mathbb{R}^N$ is related to the observations according to the model

$$\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{h}(\boldsymbol{\theta}), \boldsymbol{\Sigma}) \tag{1}$$

where $\boldsymbol{h}(\cdot) : \mathbb{R}^N \longmapsto \mathbb{R}^M$ is a known nonlinear function and $\boldsymbol{\Sigma} \in \mathbb{S}^N$ is the known noise covariance matrix. Here, the covariance matrix might be non-diagonal, i.e., the observations might be correlated across the sensors.

We acquire the data $\boldsymbol{x}$ using the *sparse sensing function* $\boldsymbol{\Phi}(\boldsymbol{w}) \in \{0,1\}^{K\times M}$ to obtain a reduced-dimensional (i.e, $K \ll M$) observation vector

$$\boldsymbol{y} = \boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{x}. \tag{2}$$

The sensing function $\boldsymbol{\Phi}(\boldsymbol{w})$ is guided by a sparse vector (thus the name sparse sensing)

$$\boldsymbol{w} = [w_1, w_2, \ldots, w_M]^T \in \{0,1\}^M.$$

Specifically, $\boldsymbol{\Phi}(\boldsymbol{w})$ is structured as

$$\boldsymbol{\Phi}(\boldsymbol{w}) = \mathrm{diag_r}(\boldsymbol{w})$$

with $\mathrm{diag_r}(\cdot)$ representing a diagonal matrix with the argument on its diagonal but with the all-zero rows removed. The sensing scheme is illustrated in Fig. 1.

The main goal is to design a sparsest $\boldsymbol{w}$ in order to guarantee a desired estimation accuracy, where the reduced-dimensional vector $\boldsymbol{y}$ is used to solve the non-linear inverse problem.

III. Inference performance metric

For parameter estimation, the estimation quality is essentially determined by the error covariance matrix $\boldsymbol{E} = \mathbb{E}\{(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^T\}$. Here, $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^N$ is an estimate of $\boldsymbol{\theta}$. Although for linear estimation problems in Gaussian noise, $\boldsymbol{E}$ admits a known expression that is independent of $\boldsymbol{\theta}$, for nonlinear models such as (1), $\boldsymbol{E}$ can only be empirically evaluated. Consequently, it is not suitable for numerical optimization. Hence, in what follows we discuss a simpler performance metric, which can be computed in closed form and is suitable for numerical optimization. More specifically, we will consider the Cramér-Rao bound (CRB) as the performance metric because (a) the CRB ensures (local) solvability of the problem, and (b) the subset of sensors that yields a lower CRB also generally yields a lower estimation error as well.

The covariance of an unbiased estimate $\widehat{\boldsymbol{\theta}}$ satisfies the following inequality [16]

$$\mathbb{E}\{(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}})^T\} \geq \boldsymbol{C}(\boldsymbol{w}, \boldsymbol{\theta}) = \boldsymbol{F}^{-1}(\boldsymbol{w}, \boldsymbol{\theta}),$$

where $\boldsymbol{C}(\boldsymbol{w}, \boldsymbol{\theta})$ is the CRB matrix and the inverse of the CRB matrix, i.e.,

$$\boldsymbol{F}(\boldsymbol{w}, \boldsymbol{\theta}) = \mathbb{E}\left\{\left(\frac{\partial \ln p(\boldsymbol{y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial \ln p(\boldsymbol{y};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)^T\right\} \in \mathbb{R}^{N\times N}$$

is the Fisher information matrix (FIM).

For the observations in (2), given (1), we can compute the FIM as

$$\boldsymbol{F}(\boldsymbol{w}, \boldsymbol{\theta}) = [\boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{J}(\boldsymbol{\theta})]^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{w}) [\boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{J}(\boldsymbol{\theta})] \tag{3}$$

where $\boldsymbol{J}(\boldsymbol{\theta}) = \partial \boldsymbol{h}(\boldsymbol{\theta})/\partial \boldsymbol{\theta}^T \in \mathbb{R}^{M\times N}$ is the Jacobian matrix and

$$\boldsymbol{\Sigma}(\boldsymbol{w}) = \boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{\Sigma}\boldsymbol{\Phi}^T(\boldsymbol{w}) \in \mathbb{R}^{K\times K}$$

is the submatrix of $\boldsymbol{\Sigma}$, which includes only the entries corresponding to the selected measurements.

A scalar function of the matrix $\boldsymbol{F}(\boldsymbol{w}, \boldsymbol{\theta})$ is optimized over all the possible samplers. In particular, some of the commonly used (in the optimal design of experiments) scalar measures are, respectively

1) $f(\boldsymbol{w}) := \mathrm{tr}\{\boldsymbol{F}^{-1}(\boldsymbol{w}, \boldsymbol{\theta})\}$,
2) $f(\boldsymbol{w}) := \lambda_{\min}\{\boldsymbol{F}(\boldsymbol{w}, \boldsymbol{\theta})\}$,
3) $f(\boldsymbol{w}) := \ln\det\{\boldsymbol{F}(\boldsymbol{w}, \boldsymbol{\theta})\}$

related to the A-, E-, and D-optimal designs. In what follows, we will derive samplers based on the E-optimal criterion due to complexity and performance reasons [3]. Nevertheless, sparse samplers can be obtained (along similar lines) by optimizing any one of the above performance metrics.

Due to the nonlinear measurement model, the FIM depends on the unknown true parameter vector $\boldsymbol{\theta}$. In practice, the unknown vector $\boldsymbol{\theta}$ has a physical meaning and takes values within a certain domain denoted by a set $\mathcal{U}$. Therefore, $f(\boldsymbol{w})$ has to be optimized over $\boldsymbol{w}$ and for all $\boldsymbol{\theta} \in \mathcal{U}$. In contrast, for linear measurement models in Gaussian noise, the FIM does not depend on the unknown parameter vector $\boldsymbol{\theta}$. Having described the inference performance metric, we now formally state the problem.

Problem statement. *Given the model parameters* $\boldsymbol{h}(\cdot) : \mathbb{R}^N \longmapsto \mathbb{R}^M$ *and* $\boldsymbol{\Sigma} \in \mathbb{R}^{M\times M}$ *in* (1)*, and a desired inference performance* λ*, design a sparsest* $\boldsymbol{w} \in \{0,1\}^M$ *that satisfies the performance constraint* $\lambda_{\min}\{\boldsymbol{F}(\boldsymbol{w}, \boldsymbol{\theta})\} \geq \lambda$, $\forall \boldsymbol{\theta} \in \mathcal{U}$.

Mathematically, this can be cast as the following optimization problem

$$\begin{aligned} \boldsymbol{w}^\star = \arg\min_{\boldsymbol{w}} \quad & \|\boldsymbol{w}\|_0 \\ \text{s.t. } & \lambda_{\min}\{\boldsymbol{F}(\boldsymbol{w}, \boldsymbol{\theta})\} \geq \lambda,\ \forall \boldsymbol{\theta} \in \mathcal{U}, \\ & \boldsymbol{w} \in \{0,1\}^M. \end{aligned} \tag{4}$$

This is a Boolean constrained cardinality minimization problem (thus nonconvex). In the next section, we will derive a solution to (4) based on some convex approximations.

IV. Proposed solver

To begin with, we relax the Boolean constraint to its best convex approximation, i.e., a box constraint $\boldsymbol{w} \in [0,1]^M$. The constraint $\lambda_{\min}\{\boldsymbol{F}(\boldsymbol{w},\boldsymbol{\theta})\} \geq \lambda$ in its current form is not convex on $\boldsymbol{w} \in [0,1]^M$. This is also true for the trace and determinant constraints.

Next, we will provide some steps to express the minimum eigenvalue constraint as a convex constraint on $\boldsymbol{w} \in [0,1]^M$. Firstly, we express the noise covariance matrix $\boldsymbol{\Sigma}$ as

$$\boldsymbol{\Sigma} = a\boldsymbol{I} + \boldsymbol{S}, \tag{5}$$

where a nonzero $a \in \mathbb{R}$ is chosen such that $\boldsymbol{S} \in \mathbb{R}^{M\times M}$ is invertible and well-conditioned. Using (5) in (3), we obtain

$$\begin{aligned}\boldsymbol{F}(\boldsymbol{w},\boldsymbol{\theta}) =& \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{\Phi}^T(\boldsymbol{w})\left(a\boldsymbol{I} + \boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{S}\boldsymbol{\Phi}^T(\boldsymbol{w})\right)^{-1} \\ &\times \boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{J}(\boldsymbol{\theta}).\end{aligned}$$

Using the fact that $\boldsymbol{\Phi}^T(\boldsymbol{w})\boldsymbol{\Phi}(\boldsymbol{w}) = \mathrm{diag}(\boldsymbol{w})$ and applying the matrix inversion lemma [16]

$$\boldsymbol{C}(\boldsymbol{B}^{-1}+\boldsymbol{C}^T\boldsymbol{A}^{-1}\boldsymbol{C})^{-1}\boldsymbol{C}^T = \boldsymbol{A} - \boldsymbol{A}(\boldsymbol{A}+\boldsymbol{C}\boldsymbol{B}\boldsymbol{C}^T)^{-1}\boldsymbol{A},$$

with $\boldsymbol{C} = \boldsymbol{\Phi}^T(\boldsymbol{w})$, $\boldsymbol{B}^{-1} = a\boldsymbol{I}$, and $\boldsymbol{A} = \boldsymbol{S}^{-1}$, it is easy to verify that

$$\begin{aligned}\boldsymbol{\Phi}^T(\boldsymbol{w})\left(a\boldsymbol{I} + \boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{S}\boldsymbol{\Phi}^T(\boldsymbol{w})\right)^{-1}\boldsymbol{\Phi}(\boldsymbol{w}) = \boldsymbol{S}^{-1} \\ - \boldsymbol{S}^{-1}\left[\boldsymbol{S}^{-1} + a^{-1}\mathrm{diag}(\boldsymbol{w})\right]^{-1}\boldsymbol{S}^{-1}.\end{aligned} \tag{6}$$

Therefore, we can simplify $\boldsymbol{F}(\boldsymbol{w},\boldsymbol{\theta})$ to

$$\begin{aligned}\boldsymbol{F}(\boldsymbol{w},\boldsymbol{\theta}) &= \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1}\boldsymbol{J}(\boldsymbol{\theta}) \\ &- \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1}\left[\boldsymbol{S}^{-1} + a^{-1}\mathrm{diag}(\boldsymbol{w})\right]^{-1}\boldsymbol{S}^{-1}\boldsymbol{J}^T(\boldsymbol{\theta}).\end{aligned} \tag{7}$$

In contrast to (3), the design parameter $\boldsymbol{w}$ appears only once in (7), which makes the problem much easier. Using the Schur complement, the constraint $\lambda_{\min}\{\boldsymbol{F}(\boldsymbol{w},\boldsymbol{\theta})\} \geq \lambda$ can now be equivalently expressed under $a > 0$ as an LMI of size $M+N$:

$$\left[\begin{array}{ll}\boldsymbol{S}^{-1} + a^{-1}\mathrm{diag}(\boldsymbol{w}) & \boldsymbol{S}^{-1}\boldsymbol{J}(\boldsymbol{\theta}) \\ \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1} & \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1}\boldsymbol{J}(\boldsymbol{\theta}) - \lambda\boldsymbol{I}_N\end{array}\right] \succeq \boldsymbol{0}_{M+N}, \tag{8}$$

which is linear (thus convex) in $\boldsymbol{w}$. Due to (8), the matrix $\boldsymbol{S}^{-1} + a^{-1}\mathrm{diag}(\boldsymbol{w})$ should be positive definite.

Note that the constraint (8) also depends on the unknown parameter vector $\boldsymbol{\theta}$. We remark here that for linear measurement models, the above constraint is independent of the unknown parameter vector $\boldsymbol{\theta}$. In other words, in that case, $\boldsymbol{J}(\boldsymbol{\theta})$ will be independent of $\boldsymbol{\theta}$ and will simply be the regression matrix itself. To arrive at (8) from the FIM in (3), is an important contribution of this work, and it is valid for arbitrary (invertible) covariance matrix $\boldsymbol{\Sigma}$.

Finally, the ℓ_0-norm cost function has to be replaced with a good convex approximation. Traditionally, the ℓ_1-norm has been used as the best convex relaxation for the ℓ_0-norm, however, it might not always yield a sparse solution (or a sparse sampler) [3]. Alternatively, a nonconvex surrogate function, namely, the sum-of-logarithms given by $\sum_{m=1}^{M}\ln(w_m+\delta)$, can yield a sparser solution due to a better approximation of the ℓ_0 norm. Although the sum-of-logarithms is concave on $\boldsymbol{w}$, since it is smooth with respect to $\boldsymbol{w}$, an iterative linearization can be performed to obtain a local minimum [3]. Specifically, we solve a re-weighted ℓ_1-norm optimization problem of the form

$$\begin{aligned}\widehat{\boldsymbol{w}}[i] = &\underset{\boldsymbol{w}\in\mathbb{R}^M}{\arg\min} \quad \sum_{m=1}^{M}\frac{w_m}{\widehat{w}_m[i-1]+\delta} \\ \text{s.t.} &\left[\begin{array}{ll}\boldsymbol{S}^{-1} + a^{-1}\mathrm{diag}(\boldsymbol{w}) & \boldsymbol{S}^{-1}\boldsymbol{J}(\boldsymbol{\theta}) \\ \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1} & \boldsymbol{J}^T(\boldsymbol{\theta})\boldsymbol{S}^{-1}\boldsymbol{J}(\boldsymbol{\theta}) - \lambda\boldsymbol{I}_N\end{array}\right] \succeq \boldsymbol{0}_{M+N}, \\ &\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \forall\boldsymbol{\theta}\in\mathcal{U}, \\ &0 \leq w_m \leq 1, \quad m = 1,\ldots,M.\end{aligned} \tag{9}$$

Here, i denotes the iteration index and $\delta > 0$ is a small constant that prevents the cost from tending to $-\infty$. Finally, the sparse sampler is given by $\widehat{\mathbf{w}} = \widehat{\mathbf{w}}[i_{\max}]$, where $i_{\max}$ is the specified maximum number of iterations. In each iteration, we solve a convex program (9), more specifically, a semidefinite program. This can be solved using any of the off-the-shelf solvers like SeDuMi [17] or YALMIP [18].

V. Numerical example

We apply the developed theory to sensor placement for the source localization setup illustrated in Fig. 2(a). In applications related to field estimation, (active/passive) radar, and sonar, it is important to estimate the location of a point source that emits or reflects energy. Suppose that there are M candidate sensors that can be placed at locations $\{\boldsymbol{a}_m \in \mathbb{R}^2\}_{m=1}^{M}$, and that measure the energy generated by a point source at location $\boldsymbol{\theta} \in \mathbb{R}^2$. The measurements are given as

$$y_m = h_m(\boldsymbol{\theta}) + n_m,\ m = 1,2,\ldots,M, \tag{10}$$

where $h_m(\cdot)$ is an isotropic exponential attenuation function given by $h_m(\boldsymbol{\theta}) = \sqrt{e}\beta/(\beta + d_m^2)$ with $d_m = \|\boldsymbol{\theta} - \boldsymbol{a}_m\|_2$. Here, e is the known field intensity emitted or reflected by the point source, $\beta \geq 0$ is the gain, and n_m is the noise. In this case, we have

$$\boldsymbol{h}(\boldsymbol{\theta}) = [h_1(\boldsymbol{\theta}), h_2(\boldsymbol{\theta}), \ldots, h_M(\boldsymbol{\theta})]^T$$

and $\boldsymbol{n} = [n_1, n_2, \ldots, n_M]^T$ with $\boldsymbol{n} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$. The noise covariance matrix $\boldsymbol{\Sigma}$ might not be diagonal due to the multipath effects, for example. Using the above model parameters, we have to choose the best subset of sensor locations out of M available locations such that a desired estimation accuracy for estimating $\boldsymbol{\theta}$ is achieved.

Consider a scenario with $M = 80$ sensors and a noise covariance matrix of the form

$$\boldsymbol{\Sigma} = \left[\begin{array}{ll}\boldsymbol{\Sigma}_{\mathrm{horz}} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_{\mathrm{vert}}\end{array}\right] \in \mathbb{R}^{M\times M},$$

where $\boldsymbol{\Sigma}_{\mathrm{horz}}$ is the noise covariance matrix corresponding to the horizontally located candidate sensors denoted by ($\lozenge$) in Fig. 2(a) and $\boldsymbol{\Sigma}_{\mathrm{vert}}$ is the noise covariance matrix corresponding to the vertically located candidate sensors indicated by ($\square$) in Fig. 2(a). We further assume that

$$\boldsymbol{\Sigma}_{\mathrm{horz}} = \sigma^2\left[(1-\rho)\boldsymbol{I} + \rho\boldsymbol{1}\boldsymbol{1}^T\right]$$

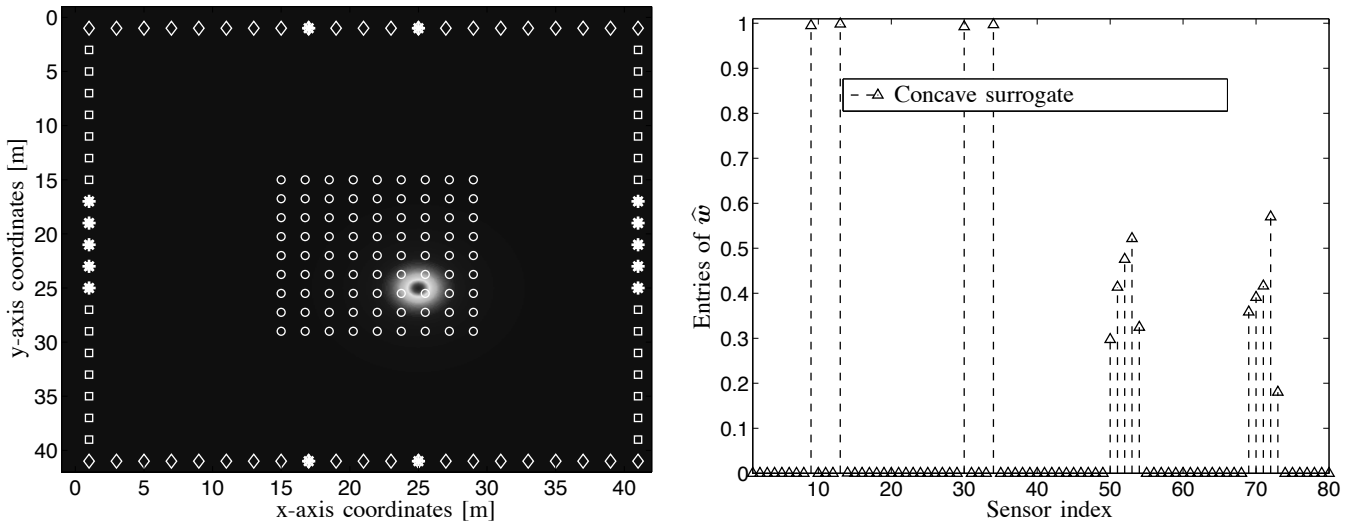

Fig. 2: Sensor placement for source localization ($\rho = 0.5$). Here, the uncorrelated and correlated sensors are denoted by squares ($\square$) and diamonds ($\lozenge$), respectively. The source domain is indicated by circles ($\circ$), while the selected sensors are indicated by ($*$). (a) Illustration of a field generated by a unit amplitude point source at location $\boldsymbol{\theta} = [25, 25]^T$ m according to (10). Out of $M = 80$ available sensors 14 sensors are selected. (b) Sensor selection is solved solved using (9) with $\lambda = 0.02$ and $i_{\max} = 10$.

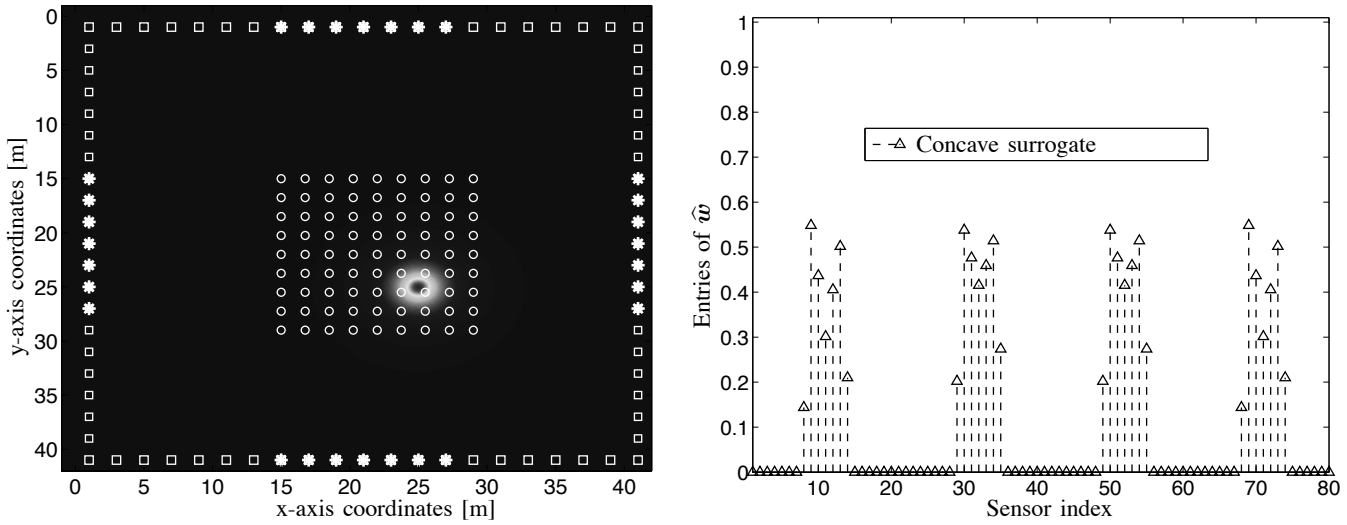

Fig. 3: Sensor placement for source localization — uncorrelated case ($\rho = 0$). Here, the candidate sensors are denoted by squares ($\square$). The source domain is indicated by circles ($\circ$), while the selected sensors are indicated by ($*$). (a) Illustration of a field generated by a unit amplitude point source at location $\boldsymbol{\theta} = [25, 25]^T$ m according to (10). Out of $M = 80$ available sensors 28 sensors are selected. (b) Sensor selection is solved using (9) with $\lambda = 0.02$ and $i_{\max} = 10$.

with correlation coefficient ρ and nominal noise variance σ^2, and $\boldsymbol{\Sigma}_{\text{vert}} = \sigma^2 \boldsymbol{I}$. That is, the vertically located candidate sensors are uncorrelated while the horizontally located candidate sensors are equally correlated. We use the following simulation parameters: $e = 1$, $\beta = 1$, $\sigma^2 = 2 \times 10^{-5}$, $\lambda = 0.02$, and $\rho = \{0, 0.5\}$.

For the correlated case with $\rho = 0.5$, the sensor placement is shown in Fig. 2(a). Here, the placement and related solution shown in Fig. 2(b) is obtained by solving (9) using SeDuMi with $i_{\max} = 10$ and $\delta = 10^{-6}$. Clearly, the obtained solution is not Boolean, and deterministic or randomized rounding [3] can be used to obtain an approximate Boolean solution. Here, we use deterministic rounding.

As a benchmark, Fig. 3(a) shows the sensor placement for the uncorrelated case, i.e., for $\rho = 0$, where the placement is again obtained by solving (9) using SeDuMi with $i_{\max} = 10$ and $\delta = 10^{-6}$, and then using deterministic rounding. Fig. 3(b) shows the solution of (9).

We underline the following observations. Firstly, the sensors from the same region are selected for both the uncorrelated and correlated case as the structure of the FIM in both these cases is similar. Secondly, to achieve the desired performance requirement we see that fewer correlated sensors are selected as compared to the number of selected uncorrelated sensors. Finally, as we observed in the simulations (not shown here), for this particular numerical example, the ℓ_1-norm (i.e., with

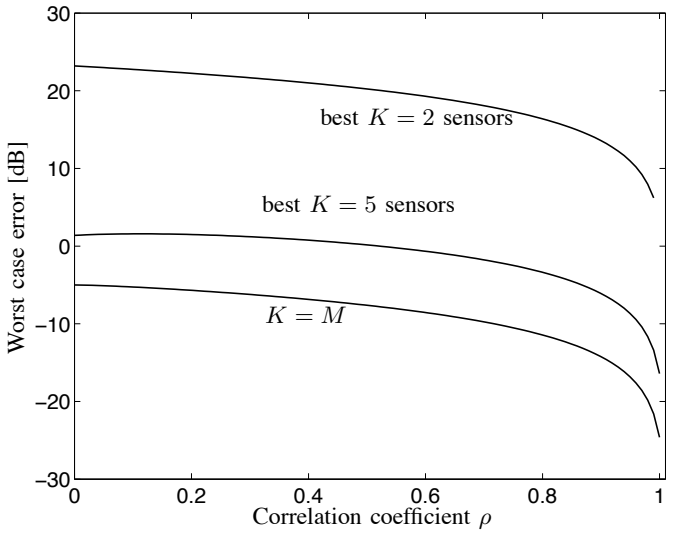

Fig. 4: Worst case error (i.e., $\lambda_{\max}\{\boldsymbol{E}(\boldsymbol{w})\}$) for different values of the correlation coefficient ρ.

$i_{\max} = 1$) based solution does not result in a sparse solution.

VI. Correlation vs. number of sensors

We will now study in more detail the impact of correlation on the number of sensors required to reach a desired inference performance. In particular, we will focus on a linear measurement model in equicorrelated Gaussian noise, i.e.,

$$\boldsymbol{x} \sim \mathcal{N}\left(\boldsymbol{H}\boldsymbol{\theta}, \boldsymbol{\Sigma}\right)$$

where

$$\boldsymbol{\Sigma} = \left[(1-\rho)\boldsymbol{I} + \rho \mathbf{1}\mathbf{1}^T\right]$$

and the regressors $\{\boldsymbol{h}_m\}_{m=1}^M$ form the rows of $\boldsymbol{H}$. This is a special case of the data model (1) for which we can compute the error covariance matrix in closed form given by

$$\boldsymbol{E}(\boldsymbol{w}) = \left(\left[\boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{H}\right]^T \boldsymbol{\Sigma}^{-1}(\boldsymbol{w}) \left[\boldsymbol{\Phi}(\boldsymbol{w})\boldsymbol{H}\right]\right)^{-1},$$

and it is equal to the Cramér-Rao bound matrix.

For independent and identically distributed Gaussian regressors $\{\boldsymbol{h}_m\}_{m=1}^M$ with $M = 10$ and $N = 2$, we show the impact of the correlation coefficient on the worst case error, i.e., $\lambda_{\max}\{\boldsymbol{E}(\boldsymbol{w})\}$, in Fig. 4. To ensure identifiability, we need $K \geq 2$. Here, the threshold λ is selected to choose the best subset of $K = \{2, 5, M\}$ sensors by solving (9). It can be observed that as the sensors become more coherent, i.e., as $\rho \to 1$, the worst case error drops significantly. In other words, it implies that the number of sensors required to reach a desired estimation accuracy reduces significantly as the sensors become more coherent. This result also holds for the mean squared error (i.e., A-optimal design), however it is not show here.

VII. Conclusions

In this paper, we have discussed discrete sparse sensing for non-linear parameter estimation problems. In particular, we design a sparse sensing function that is guided by a sparse vector to acquire colored Gaussian observations such that a desired estimation accuracy is guaranteed. We optimize scalar functions of the Cramér-Rao bound matrix, which we use as the inference performance metric to design the samplers. The sampler design does not require actual measurements, however it needs the model parameters to be known perfectly (thus can be designed offline). We derive a convex program to compute the sparse sampler of interest. To do so, we propose a decomposition of the noise covariance matrix, which is valid for arbitrary correlation values. The developed theory can be applied to sensor selection and sensor placement problems. The proposed approach is illustrated with a sensor placement example.

References

[1] 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.

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

[3] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 684–698, Feb. 2015.

[4] E. Masazade, M. Fardad, and P.K. Varshney, "Sparsity-promoting extended kalman filtering for target tracking in wireless sensor networks," *IEEE Signal Process. Lett.*, vol. 19, no. 12, pp. 845–848, 2012.

[5] F. Lin, M. Fardad, and M. R. Jovanović, "Algorithms for leader selection in stochastically forced consensus networks," *IEEE Trans. on Automatic Control*, vol. 59, no. 7, pp. 1789–1802, July 2014.

[6] S. P. Chepuri and G. Leus, "Continuous sensor placement," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 544–548, May 2015.

[7] 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.

[8] S. Liu, M. Fardad, P. K. Varshney, and E. Masazade, "Optimal periodic sensor scheduling in networks of dynamical systems," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3055–3068, June 2014.

[9] V. Roy and G. Leus, "Correlation-aware sparsity-enforcing sensor placement for spatio-temporal field estimation," in *Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2015, pp. 2389–2393.

[10] V. Roy, S.P. Chepuri, and G. Leus, "Sparsity-enforcing sensor selection for DOA estimation," in *Proc. of the IEEE Workshop on Comp. Adv. in Multi-Sensor Adaptive Proc. (CAMSAP)*, Dec. 2013.

[11] 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.

[12] S. P. Chepuri and G. Leus, "Sparse sensing for distributed gaussian detection," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2015.

[13] S.P. Chepuri and G. Leus, "Sparse sensing for distributed detection," *IEEE Trans. on Sig. Process.*, 2015.

[14] E. Rigtorp, "Sensor selection with correlated noise," M.S. thesis, KTH Royal Institute of Technology, Aug. 2010.

[15] H.-J. Rad, A. Simonetto, X. Ma, and G. Leus, "Sparsity-aware sensor selection for correlated noise," in *Proc. of Intl. Conf. on Information Fusion (FUSION), Spain*, Jul. 2014.

[16] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*, Englewood Cliffs, NJ: Prentice-Hall, 1993.

[17] 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.

[18] J. Lofberg, "Yalmip : A toolbox for modeling and optimization in MATLAB," in *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.