

Sensor Array Design Through Submodular Optimization

Gal Shulkind¹, Student Member, IEEE, Stefanie Jegelka, and Gregory W. Wornell², Fellow, IEEE

Abstract—We consider the problem of far-field sensing by means of a sensor array. Traditional array geometry design techniques are agnostic to prior information about the far-field scene. However, in many applications such priors are available and may be utilized to design more efficient array topologies. We formulate the problem of array geometry design with scene prior as one of finding a sampling configuration that enables efficient inference, which turns out to be a combinatorial optimization problem. While generic combinatorial optimization problems are NP-hard and resist efficient solvers, we show how for array design problems the theory of submodular optimization may be utilized to obtain efficient algorithms that are guaranteed to achieve solutions within a constant approximation factor from the optimum. We leverage the connection between array design problems and submodular optimization and port several results of interest. We demonstrate efficient methods for designing arrays with constraints on the sensing aperture, as well as arrays respecting combinatorial placement constraints. This novel connection between array design and submodularity suggests the possibility for utilizing other insights and techniques from the growing body of literature on submodular optimization in the field of array design.

Index Terms—Array design, submodular optimization, far field sensing.

I. INTRODUCTION

SENSOR arrays for spatial sensing are widely deployed in a wide range of applications including radar, sonar, medical imaging and radio astronomy and there is a vast literature on the topics of array design and array processing from the last century [1], [2].

A major goal in designing arrays is efficiently meeting sensing specifications with a limited budget of sensing elements, which is often a main determinant of system cost in terms of dimensions, weight, complexity and manufacturing cost. However, traditionally most studies assume a uniform and linear design, often a truncated half-wavelength array, or

restrictions thereof. Indeed, the problem of designing non-uniform arrays hints at combinatorial optimization and is computationally intractable as we discuss later.

In beamforming arrays attaining a desired resolution level is often a primary concern, achieved by means of manipulating beam pattern parameters such as lobe widths and positions. The problem of designing efficient array geometries that facilitate desired beam patterns has been widely studied in the past. Techniques involving array thinning start with a dense uniform geometry, removing elements while maintaining performance within specified bounds [3]. Other approaches consider various methods such as swarm optimization [4], dynamic programming [5], genetic algorithms [6], inversion [7] and Bayesian compressive sampling [8].

In direction-of-arrival estimation applications other specialized techniques have emerged for finding efficient array designs, such as optimizing the corresponding Cramer-Rao error bound [9], or designing according to the nested array methodology for increasing the available number of degrees of freedom [10], and theoretical results for estimation performance in these settings have been developed [11], [12], [13]. Alternatively, other studies have addressed array design from the point of view of performing efficient compressive sampling of a Wide Sense Stationary (WSS) process, e.g. for a setting with a parametrized covariance matrix [14], for wide-band power spectrum estimation [15] and evaluating performance in the presence of modeling errors [16]. However, these samplers are optimized for reconstructing the second order statistics of the scene and not the actual scene realization.

In virtually all design methodologies some assumptions are made on the scene of interest. These represent beliefs, constraints or knowledge that hold over the unobserved scene. A favorable design is one meeting requirements while taking into account these assumptions. For example, in direction of arrival applications we may assume some limit on the number of point targets present in the scene [9]. In beamforming we assume some separation level between objects of interest that implies a certain resolution level is required [1], or some scene sparsity structure [17], [18].

In this paper we study the problem of inference on a scene of interest through measurements collected by an array of sensors. The approach we take for designing array geometries is novel in that we consider settings where some Bayesian prior on the scene is available at the time of design and propose exploiting this knowledge by adapting the array geometry

Manuscript received January 19, 2017; revised December 28, 2017; accepted September 16, 2018. Date of publication October 5, 2018; date of current version December 19, 2018. This work was supported in part by the NSF CAREER Award 1553284, in part by DARPA under Contract HR0011-16-C-0030, and in part by NSF under Grant CCF-1717610. Some part of the work in this paper was presented at the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing.

The authors are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mail: fshulkind@mit.edu; stefje@mit.edu; gwwg@mit.edu).

Communicated by R. La, Associate Editor for Communication Networks.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIT.2018.2873795

accordingly to achieve efficient inference with a limited budget for sensors.

Prior knowledge is in fact not a rarity at all. Frequently, the same device is used to sense multiple similar scenes, where past examples are indicative. A medical imaging device, for instance, is typically used to image the same organ across different patients. In other cases, we may have prior knowledge in the form of scene properties such as smoothness or adherence to spatial constraints. We incorporate such knowledge as a prior in a Bayesian model, such that sensing the scene is equivalent to performing inference in this model, and the problem of array geometry design asks to select a geometry that optimizes the quality of inference. This naturally turns out to be a combinatorial optimization problem, which can be extremely hard to solve even approximately without exploiting additional structure.

We show how to choose a cost function for the array geometry design problem that holds the property of submodularity [19]. A submodular set function is one that exhibits diminishing marginal gains, i.e. adding additional elements results in diminishing benefit. Recently, there has been significant progress on the theory of optimizing submodular functions [20]–[23]. In particular, these results state that, while NP-hard, submodular maximization admits variants of greedy algorithms that are guaranteed to achieve near-optimal solutions, i.e. within a constant factor. Submodularity has been used in connection with sensor placement problems, for example in [24]–[26], however these works are not tailored to the far field scenes and models that we focus on here.

Our connections and formulations open avenues for leveraging those results for efficient array design with strong guarantees. We demonstrate this by showcasing the design of array geometries in settings with arbitrary apertures and settings where sensor placement respects matroid combinatorial constraints by porting results from the growing body of literature on submodular optimization. We demonstrate that together with our new adaptive formulations, the exploitation of prior knowledge leads to higher quality inference at lower cost in terms of the number of sensing elements.

The paper is structured as follows: In Section II, we review far-field sensing and introduce a statistical framework for modeling a-priori scene distributions. In Section III, we define a cost-function to quantify the inference loss associated with array configurations and derive a corresponding array design optimization problem. We show that this optimization problem belongs to the class of submodular optimization problems and adapt established algorithms and guarantees to our setting. In Section IV, we review the theory of matroids and showcase its use in designing array geometries with combinatorial placement constraints. Finally, in Section V we summarize numerical experiments exemplifying and validating our theory. The appendix provides proofs for some of our claims.

II. PROBLEM FORMULATION

In this section we formulate the array design problem. We focus on far-field sensing applications, although, as will become apparent, the same techniques could also be generalized to other settings where the measurement process

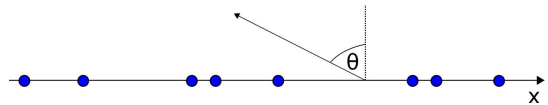


Fig. 1. Far-field sensing. Antenna elements are depicted as blue dots.

is linear. For simplicity we consider a one dimensional setting as the extension to multiple dimensions is straightforward. We begin with a review of the far-field sensing model to establish notation, and then pose the sensing problem as a Bayesian inference problem.

A. Far-Field Sensing

The far-field sensing setup is depicted in Fig. 1. At a distance from an observation axis x we have a scene of interest, characterized by an illumination function $\tilde{\beta}(\theta)$ that depends on the angle $\theta \in [-\frac{\pi}{2}, +\frac{\pi}{2}]$ between the direction of observation and the axis of measurement broadside. We are to place N sensors along the observation axis at positions $\mathcal{S} \equiv \{x_1, \dots, x_N\}$.

The passive sensors pick up the radiation emitted from the scene, which is assumed to be narrow-band around a fixed wavelength λ . Recall that in an ideal noiseless setting, $r(x)$, the reading collected at position x , is given according to [1]:

$$r(x) = \int_{-\frac{\pi}{2}}^{+\frac{\pi}{2}} \tilde{\beta}(\theta) e^{j\frac{2\pi}{\lambda}x \sin(\theta)} \cos(\theta) d\theta = \int_{-\frac{1}{2}}^{+\frac{1}{2}} \beta(\psi) e^{j\frac{4\pi}{\lambda}x\psi} d\psi, \quad (1)$$

where we have introduced $\psi \equiv \frac{\sin\theta}{2}$ such that $\psi \in [-\frac{1}{2}, +\frac{1}{2}]$ and $\beta(\psi) \equiv \tilde{\beta}(\sin^{-1}(2\psi))$.

Eq. (1) bears close resemblance to the definition of the Fourier transform. Namely, defining $f(t)$ to be the Fourier transform pair of $\beta(\psi)$

$$f(t) = \int \beta(\psi) e^{j2\pi t\psi} d\psi \xrightarrow[\mathcal{F}^{-1}]{\mathcal{F}} \beta(\psi) = \int f(t) e^{-j2\pi t\psi} dt \quad (2)$$

we immediately identify, comparing (1) and (2) that $r(x) = f(\frac{2}{\lambda}x)$. That is, collecting measurements at $x \in \{x_n\}$ is equivalent to sampling the function $f(t)$ at $t \in \{\frac{2}{\lambda}x_n\}$. Notice that since the support of $\beta(\psi)$ is restricted to $\psi \in [-\frac{1}{2}, +\frac{1}{2}]$ we observe that $f(t)$ is band-limited with bandwidth 1. Thus, as a consequence of the Whittaker-Kotelnikov-Shannon sampling theorem we have that perfect reconstruction of $f(t)$, and subsequently the scene $\beta(\psi)$, is possible from an infinite set of samples taken at $t \in \{\dots, -1, 0, +1, \dots\}$ which are exactly the samples collected by an infinite $\frac{1}{2}$ -spaced array.

Finally, we consider the effect of noise by introducing the vector of N noisy measurements $\tilde{\mathbf{f}} = [\tilde{f}_1, \dots, \tilde{f}_N]^T$ modelled according to:

$$\tilde{f}_n \equiv r(x_n) + w_n = f\left(\frac{2}{\lambda}x_n\right) + w_n \quad n = 1, \dots, N \quad (3)$$

where w_n are additive noise components. Stacked into a vector $\mathbf{w} = [w_1, \dots, w_N]^T$, we assume throughout that the

noise is a complex, circular, Gaussian random vector $\mathbf{w} \sim \mathcal{CN}(0, \Sigma_{ww})$ [27]. We assume the noise to be independent across different sensors, i.e., $\Sigma_{ww} = \sigma_w^2 \mathbf{I}_N$, where \mathbf{I}_N is an $N \times N$ identity matrix such that the noise is i.i.d. across different sensors.

B. Bayesian Formulation

The sensing problem we consider in this paper entails the estimation of the illumination function $\beta(\psi)$ or equivalently its inverse Fourier transform $f(t)$ from the set of noisy measurements $\tilde{\mathbf{f}}$. Even in the noiseless setting this problem is gravely ill-posed as infinitely many wildly varying scenes map to any given finite set of observations¹.

To cope with this ill-posedness, some prior belief or knowledge pertaining to $\beta(\psi)$ (or equivalently $f(t)$) must hence be incorporated into the model, and this could be achieved in several ways. Wingham [29] proposed selecting one specific $f(t)$ of the multiple such functions consistent with the samples, namely the minimum norm solution. Alternatively, some constraints or other preferences may be imposed on the solution by penalizing the inversion. For example one may require the solution to satisfy some constraints, e.g. lie in some pre-specified sub-space of the function space, or regularize the inversion, e.g. to promote smoothness or low total variation [30].

In this study we take a Bayesian approach and impose a prior distribution on the scene $\beta(\psi)$. Subsequently, sensing is equivalent to performing inference in this model. The prior may be assigned based on past observations over the distribution of scenes or based on a-priori knowledge of scene properties.

We conceptually think of our scene as a natural image, and as such, our Bayesian formulation entails posing a prior that is consistent with its expected characteristics, tailored to the specific application that is of interest to the user. Utilizing Gaussian distributions as priors for natural scenes and images has been successfully applied in the past [31], [32], [33] and our work uses similar notions, with the Gaussian prior specifically expressed in the frequency domain (e.g., as in [9]). We note that this Gaussian description naturally captures the statistics of spread-out scenes and is not immediately suitable for serving as a prior for point-source target models. The formulations and methods that we develop in subsequent sections can be adapted for use with other scene priors in lieu of the Gaussian one we focus on for this study. In particular, our formulations may account for settings with point-source scenes by posing a point-process prior (e.g. a determinantal point process, DPP [34]) that is better suited for capturing priors of point-targets in specific environments. With this alternative prior we could follow similar steps to develop efficient array designs for those applications. Studying such alternative priors and resulting formulations is a possible interesting extension to our work.

In what follows, to be concrete we detail one specific choice for the form of the scene prior, expressed as a Gaussian distribution in the frequency domain:

¹The mapping between a band-limited function $f(t)$ and a finite set of its (generally non-uniform) samples $\{f(t_i)\}$ is not bijective [28].

1) *Discrete Representation:* Assigning a prior on a continuous function $\beta(\psi)$ may seem like a daunting task. However, in many scenarios this can be considerably simplified if the scene may be efficiently expanded in a countable basis of functions such that the prior may be imposed in the discrete domain of expansion coefficients. In what follows, we consider expanding $\beta(\psi)$ by means of Fourier basis functions $\{e^{j2\pi m\psi} | m = \dots, -1, 0, +1, \dots\}$ in $\psi \in [-\frac{1}{2}, +\frac{1}{2}]$. Alternatively, other expansions may be used with similar results. We write:

$$\beta(\psi) = \sum_m \beta_m e^{j2\pi m\psi}, \quad \psi \in [-\frac{1}{2}, +\frac{1}{2}] \quad (4)$$

$$\beta_m \equiv \int_{-\frac{1}{2}}^{+\frac{1}{2}} \beta(\psi) e^{-j2\pi m\psi} d\psi \quad (5)$$

where $\{\beta_m\}$ are the Fourier expansion coefficients. The usual Parseval relation holds:

$$\int |\beta(\psi)|^2 d\psi = \sum_m |\beta_m|^2 \quad (6)$$

In lieu of the prior on $\beta(\psi)$ we will impose a prior distribution over $\{\beta_m\}$. As will become evident such a description is especially suited for applications involving real world smooth scenes $\beta(\psi)$ as suggested by the following properties of the Fourier series expansion [35]:

- (a) (Riemann-Lebesgue lemma) Let $\beta(\psi)$ be any integrable function. Then $|\beta_m| \xrightarrow{|m| \rightarrow \infty} 0$.
- (b) Let $\beta(\psi) \in C^r$ where C^r is the space of r -times continuously differentiable functions over some domain. Then $|\beta_m| \leq \frac{\alpha}{|m|^r}$ with $\alpha = \sup_{\psi} |\frac{\partial^r}{\partial \psi^r} \beta(\psi)|$ [35].

Thus, for any nicely behaved scene $\beta(\psi)$ the high frequency Fourier expansion coefficients diminish to zero with an asymptotically polynomial rate determined by the level of smoothness.

2) *Prior:* In the sequel we use Gaussian priors on the coefficients $\{\beta_m\}$:

$$\beta_m \sim \mathcal{CN}(0, \sigma_m^2) \quad (7)$$

where β_m are independent, complex, circular and Gaussian, and σ_m^2 are the corresponding variances, which are assumed known in advance. Using (6) we define the expected scene power:

$$P \equiv \mathbb{E} \int |\beta(\psi)|^2 d\psi = \sum_m \sigma_m^2 \quad (8)$$

The $\{\sigma_m^2\}$ can be set following some initial measurements of sample functions $\beta(\psi)$ or taking into account prior knowledge. For example if we have a-priori knowledge that $\beta(\psi) \in C^r$ we may use

$$\sigma_m^2 = \begin{cases} \alpha & m = 0 \\ \frac{\alpha}{m^{2r}} & m \neq 0 \end{cases} \quad (9)$$

with α a scaling factor which is a very simple distribution respecting the polynomial variance decay. For the rest of the paper we adopt the prior in (9) and take $\alpha = 1$ and $r = 1$

to promote continuously differentiable functions. Also notice that the resulting prior distribution on $\beta(\psi)$ is Wide Sense Stationary (WSS), with the coefficients σ_m^2 determining the shape of the power spectrum, based on the available prior information we have on the scene.

C. Observation Model

With the prior stated in the discrete $\{\beta_m\}$ domain as described above, our next goal is to circumvent $\beta(\psi)$ by directly stating the problem in terms of the measurements \tilde{f}_n and the coefficients β_m , replacing the continuous representation with manageable discrete counterparts. Substituting the sum (4) into Equation (1) we have:

$$r(x_n) = \int_{-\frac{1}{2}}^{+\frac{1}{2}} \sum_m \beta_m e^{j2\pi m\psi} e^{j\frac{4\pi}{\lambda} x_n \psi} d\psi = \sum_m K_{nm} \beta_m \quad (10)$$

where

$$K_{nm} \equiv \int_{-\frac{1}{2}}^{+\frac{1}{2}} e^{j2\pi(\frac{2}{\lambda}x_n+m)\psi} d\psi = \text{sinc}(m + \frac{2}{\lambda}x_n) \quad (11)$$

with $\text{sinc}(x) \equiv \frac{\sin(\pi x)}{\pi x}$. Plugging this into (3) we finally have the full observation model

$$\tilde{f}_n = \sum_m K_{nm} \beta_m + w_n \quad n = 0, \dots, N-1 \quad (12)$$

and the sensing problem amounts to estimating the coefficients $\{\beta_m\}$ given the observation vector $\tilde{\mathbf{f}}$. As we have assumed a Gaussian distribution for the noise vector \mathbf{w} as well as for the prior over $\{\beta_m\}$ the posterior distribution $p(\{\beta_m\}|\tilde{\mathbf{f}})$ turns out to be Gaussian with a convenient analytic expression, as we detail in Section III-D.

III. ARRAY DESIGN WITH AN APERTURE CONSTRAINT

In the previous section we formulated the problem of far-field sensing in a Bayesian setting with a prior on the distribution of the underlying scene. In this section we design corresponding array geometries to facilitate efficient sensing that exploit the prior knowledge and model. Initially we consider simple constraints on the total number of sensors we can place and on the aperture where these can be situated. Specifically, we assume that up to N sensors are free to be placed over some aperture \mathcal{A} on the real line, e.g. for simplicity we can consider $\mathcal{A} \equiv \{x | -a \leq x \leq a\}$, $a \in \mathbb{R}^+$. In Section IV we consider more sophisticated combinatorial placement constraints and show that the same formulations we develop here may be adapted to those more challenging use cases.

A. Formulating a Cost Function

In order to make the problem of optimal array design well-posed in our Bayesian setting we need to specify a performance measure that will be used as a cost function to

compare different designs and choose the best one. Revisiting our observation model (12) we notice that the array design determines the coefficients K_{nm} through the set of sensor positions \mathcal{S} , as defined in Section II-A. With any given design the sensing problem amounts to inferring the posterior $p(\{\beta_m\}|\tilde{\mathbf{f}})$. A natural performance measure in this setting quantifies the quality of inference, i.e., the information gained by performing the sensing experiments which results in updating our beliefs about the coefficients $\{\beta_m\}$ from the prior $p(\{\beta_m\})$ to the posterior $p(\{\beta_m\}|\tilde{\mathbf{f}})$. This problem has been extensively studied in the context of statistical inference and experimental design [36], [37]. If we are interested in making general inference to learn the state of the world (represented by the posterior distribution of the Fourier coefficients) then the natural performance measure, which is referred to as Bayes D-optimality is the mutual information between the expansion coefficients and the collected measurements [36], [37]:

$$I(\tilde{\mathbf{f}}_{\mathcal{S}}; \{\beta_m\}) = H(\{\beta_m\}) - H(\{\beta_m\}|\tilde{\mathbf{f}}_{\mathcal{S}}) \quad (13)$$

where $I(\cdot; \cdot)$ is the mutual information and $H(\cdot)$ the Shannon entropy. The subscript \mathcal{S} explicitly emphasizes the dependence of the measurements on the set of sensor positions \mathcal{S} . Notice that this cost function measures the reduction in uncertainty of the scene expansion coefficients before and after measurements are made. The larger the mutual information the more we trust the values of the coefficients $\{\beta_m\}|\tilde{\mathbf{f}}_{\mathcal{S}}$.

With the cost function in place the array design problem becomes

$$S^* = \underset{S \subseteq \mathcal{A}, |S| \leq N}{\text{argmax}} I(\tilde{\mathbf{f}}_{\mathcal{S}}; \{\beta_m\}), \quad (14)$$

which is an NP-hard combinatorial optimization problem. However, we will show later that we can obtain a constant-factor approximation for (14) using efficient computational techniques.

B. Finite Dimensional Approximation

Solving (14) under our model (12) involves manipulations of the infinite sequence $\{\beta_m\}$ which may not be amenable to computer representation. To make our formulation tractable we approximate the infinite sequence $\{\beta_m\}$ with a finite truncated set of coefficients $\{\beta_m | m \in \mathcal{M}\}$, where \mathcal{M} is some finite set. Stacked in vector form $\boldsymbol{\beta}$, we consider the simplified finite dimensional approximation of (12):

$$\hat{\mathbf{f}}_{\mathcal{S}} = \mathbf{K}_{\mathcal{S}} \boldsymbol{\beta} + \mathbf{w} \quad (15)$$

where $\mathbf{K}_{\mathcal{S}}$ is an $N \times |\mathcal{M}|$ matrix formed by restricting K_{nm} on $m \in \mathcal{M}$ and the dependency on the sampling set \mathcal{S} again made explicit. Notice that hat notation is replacing the previous tilde.

We show that the approximate finite-dimensional model (15) is a good proxy for the original infinite-dimensional model (12) for a suitably selected \mathcal{M} . Indeed, if \mathcal{M} is chosen to only exclude those β_m coefficients that in expectation contribute a marginally small part of the energy of the full infinite sequence $\{\beta_m\}$ then the mutual information derived from the approximate model (15) will closely track that of the infinite dimensional model in (12). More precisely we have the following result:

Lemma 1: Let the prior on $\{\beta_m\}$ be i.i.d. according to $\beta_m \sim \mathcal{CN}(0, \sigma_m^2)$, \mathcal{M} fixed, and $\epsilon \equiv \sum_{m \notin \mathcal{M}} \sigma_m^2$ satisfies $\epsilon < \sigma_w^2 N^{-\frac{3}{2}}$.

We have:

$$\begin{aligned} -N \log\left(1 + \frac{\epsilon N^{\frac{3}{2}}}{\sigma_w^2}\right) &\leq I(\tilde{\mathbf{f}}_{\mathcal{S}}; \{\beta_m\}) - I(\hat{\mathbf{f}}_{\mathcal{S}}; \boldsymbol{\beta}) \\ &\leq -N \log\left(1 - \frac{\epsilon N^{\frac{3}{2}}}{\sigma_w^2}\right) \end{aligned}$$

Proof: See Appendix A. \square

By virtue of the last lemma and $N \log(1 \pm \frac{\epsilon N^{\frac{3}{2}}}{\sigma_w^2}) \xrightarrow{\epsilon \rightarrow 0} 0$

we have that $I(\hat{\mathbf{f}}_{\mathcal{S}}; \boldsymbol{\beta})$ is an arbitrarily accurate proxy for $I(\tilde{\mathbf{f}}_{\mathcal{S}}; \{\beta_m\})$ for ϵ small enough, such that in lieu of problem (14) we now continue with the simplified finite dimensional approximation:

$$\mathcal{S}^* = \underset{\mathcal{S} \subseteq \mathcal{A}, |\mathcal{S}| \leq N}{\operatorname{argmax}} I(\hat{\mathbf{f}}_{\mathcal{S}}; \boldsymbol{\beta}) \quad (16)$$

and the results will be accurate to within the approximation bounds from Lemma 1.

C. Grid Discretization

Next we turn to discretizing the aperture \mathcal{A} to cast the array design problem in the form of a generic finite selection problem. We thus restrict the choice of sampling positions to the finite set

$$\mathcal{V} \equiv \{x^1, \dots, x^{|\mathcal{V}|}\} \subset \mathcal{A} \quad (17)$$

For the sequel we take \mathcal{V} to be a uniform δ -spaced grid of positions in \mathcal{A} . We adapt the array design problem (16) accordingly as:

$$\mathcal{S}_d^* = \underset{\mathcal{S} \subseteq \mathcal{V}, |\mathcal{S}| \leq N}{\operatorname{argmax}} I(\hat{\mathbf{f}}_{\mathcal{S}}; \boldsymbol{\beta}) \quad (18)$$

with the subscript d implying discretization. The next result can be used to quantify the level of discretization δ necessary to guarantee performance close to optimum within some specified error bound:

Lemma 2: With \mathcal{V} a uniform grid of sampling positions with distance δ between adjacent positions we have:

$$\begin{aligned} I(\hat{\mathbf{f}}_{\mathcal{S}_d^*}; \boldsymbol{\beta}) &\leq I(\hat{\mathbf{f}}_{\mathcal{S}^*}; \boldsymbol{\beta}) \leq I(\tilde{\mathbf{f}}_{\mathcal{S}^*}; \boldsymbol{\beta}) \\ &\quad + N \log\left(1 + \frac{4\delta P(1 + \frac{\delta}{\lambda})N^{\frac{3}{2}}}{\lambda\sigma_w^2}\right) \end{aligned}$$

Proof: See Appendix B. \square

With this last lemma in place the array design problem (16) may be further approximated in the more convenient finite combinatorial problem form of (18) with guarantees on the accuracy of the resulting designs. In the sequel we assume that δ is chosen such as to meet desired accuracy levels, as prescribed in Lemma 2, and work with the simplified formulation (18).

D. Sensing

With the observation model of (15), coupled with a Gaussian distribution for the noise vector \mathbf{w} and a Gaussian prior for the coefficients vector $\boldsymbol{\beta}$, calculation of the posterior distribution $\boldsymbol{\beta} | \hat{\mathbf{f}}_{\mathcal{S}}$ is particularly simple and can be performed analytically. Concretely, we have as a result of all random variables being Gaussian $\boldsymbol{\beta} | \hat{\mathbf{f}}_{\mathcal{S}} \sim \mathcal{CN}(\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\Sigma}})$ and the parameters are given according to the conventional Gaussian conditional parameters:

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \boldsymbol{\Sigma}_{\beta\hat{\mathbf{f}}} \boldsymbol{\Sigma}_{\hat{\mathbf{f}}\hat{\mathbf{f}}}^{-1} \hat{\mathbf{f}}_{\mathcal{S}} \\ \hat{\boldsymbol{\Sigma}} &= \boldsymbol{\Sigma}_{\beta\beta} - \boldsymbol{\Sigma}_{\beta\hat{\mathbf{f}}} \boldsymbol{\Sigma}_{\hat{\mathbf{f}}\hat{\mathbf{f}}}^{-1} \boldsymbol{\Sigma}_{\hat{\mathbf{f}}\beta}^\dagger \end{aligned} \quad (19)$$

where $\boldsymbol{\Sigma}_{\beta\hat{\mathbf{f}}} = \boldsymbol{\Sigma}_{\beta\beta} \mathbf{K}_{\mathcal{S}}^\dagger$, $\boldsymbol{\Sigma}_{\hat{\mathbf{f}}\hat{\mathbf{f}}} = \mathbf{K}_{\mathcal{S}} \boldsymbol{\Sigma}_{\beta\beta} \mathbf{K}_{\mathcal{S}}^\dagger + \boldsymbol{\Sigma}_{ww}$ and $\boldsymbol{\Sigma}_{\beta\beta} = \operatorname{diag}[\sigma_1^2, \dots, \sigma_M^2]$.

E. Submodular Optimization

The next step is to prescribe an efficient algorithm for the solution of (18). As we will show shortly (18) is an instance of a known NP-hard problem such that it is widely believed that no efficient algorithm for its solution exists. However, due to the structure of the cost function an efficient approximation algorithm is known to exist with strong theoretical guarantees. In this subsection we survey the relevant results and adapt them to our needs.

1) *Submodularity:* We begin by invoking the submodularity property of set functions [19]:

Definition 3: Let $G : 2^{\mathcal{V}} \rightarrow \mathbb{R}$ be a set function.

(a) G is **submodular** if it satisfies the property of decreasing marginals: $\forall \mathcal{S}, \mathcal{T} \subseteq \mathcal{V}$ such that $\mathcal{S} \subseteq \mathcal{T}$ and $x \in \mathcal{V} \setminus \mathcal{T}$ it holds that $G(\mathcal{S} \cup \{x\}) - G(\mathcal{S}) \geq G(\mathcal{T} \cup \{x\}) - G(\mathcal{T})$.

(b) G is **monotonic** (increasing) if $\forall \mathcal{S}, \mathcal{T} \subseteq \mathcal{V}$ s.t. $\mathcal{S} \subseteq \mathcal{T}$ we have $G(\mathcal{S}) \leq G(\mathcal{T})$.

As it turns out, our cost function is monotonic and submodular as the next result shows (this is almost identical to [38, Corollary 4]. We reproduce it here with adaptations to our setting for completeness):

Lemma 4: Let \mathcal{V} be defined as before, and define the set function $G : 2^{\mathcal{V}} \rightarrow \mathbb{R}$ according to $G(\mathcal{S}) = I(\hat{\mathbf{f}}_{\mathcal{S}}; \boldsymbol{\beta})$. Then G is submodular and monotonic (increasing).

Proof: Expanding the mutual information according to $I(x; y) = H(x) - H(x|y)$ we have:

$$\begin{aligned} G(\mathcal{S} \cup \{x\}) - G(\mathcal{S}) &= H(\hat{\mathbf{f}}_{\mathcal{S} \cup \{x\}}) - H(\hat{\mathbf{f}}_{\mathcal{S}}) \\ &\quad - [H(\hat{\mathbf{f}}_{\mathcal{S} \cup \{x\}} | \boldsymbol{\beta}) - H(\hat{\mathbf{f}}_{\mathcal{S}} | \boldsymbol{\beta})] = H(\hat{\mathbf{f}}_{\{x\}} | \hat{\mathbf{f}}_{\mathcal{S}}) - H(\hat{\mathbf{f}}_{\{x\}} | \boldsymbol{\beta}) \end{aligned} \quad (20)$$

where in the last equality we used the conditional independence of the components of $\hat{\mathbf{f}}_{\mathcal{S} \cup \{x\}}$ given $\boldsymbol{\beta}$. Substituting \mathcal{T} for \mathcal{S} we immediately get:

$$\begin{aligned} [G(\mathcal{S} \cup \{x\}) - G(\mathcal{S})] - [G(\mathcal{T} \cup \{x\}) - G(\mathcal{T})] &= H(\hat{\mathbf{f}}_{\{x\}} | \hat{\mathbf{f}}_{\mathcal{S}}) - H(\hat{\mathbf{f}}_{\{x\}} | \hat{\mathbf{f}}_{\mathcal{T}}) \end{aligned} \quad (21)$$

Using $\mathcal{S} \subseteq \mathcal{T}$ we have $H(\hat{\mathbf{f}}_{\{x\}} | \hat{\mathbf{f}}_{\mathcal{S}}) \geq H(\hat{\mathbf{f}}_{\{x\}} | \hat{\mathbf{f}}_{\mathcal{T}})$ such that $G(\mathcal{S} \cup \{x\}) - G(\mathcal{S}) \geq G(\mathcal{T} \cup \{x\}) - G(\mathcal{T})$ and G is submodular.

To prove monotonicity it is enough to show $G(\mathcal{S} \cup \{x\}) - G(\mathcal{S}) \geq 0$. This time expand the mutual information according to $I(x; y) = H(y) - H(y|x)$:

$$G(\mathcal{S} \cup \{x\}) - G(\mathcal{S}) = H(\beta|\hat{f}_{\mathcal{S}}) - H(\beta|\hat{f}_{\mathcal{S} \cup \{x\}}) \quad (22)$$

Conditioning can never increase entropy so $H(\beta|\hat{f}_{\mathcal{S}}) \geq H(\beta|\hat{f}_{\mathcal{S} \cup \{x\}})$ and the result follows. \square

2) *Efficient Solvers*: Our optimization problem (18) is the maximization of a monotonic submodular function. A greedy algorithm, shown as Algorithm 1, solves this problem to within the best possible approximation factor, as the following fundamental theorem states.

Algorithm 1 Greedy Submodular Maximization

```

1: function GREEDYMAX( $G(\cdot), \mathcal{V}, N$ )
2:    $\mathcal{S} \leftarrow \emptyset$ 
3:   for  $i = 1$  to  $N$  do
4:      $x^* = \operatorname{argmax}_{x \in \mathcal{V} \setminus \mathcal{S}} G(\mathcal{S} \cup \{x\})$ 
5:      $\mathcal{S} \leftarrow \mathcal{S} \cup \{x^*\}$ 
6:   Return  $\mathcal{S}$ 

```

The following theorem guarantees that the greedy algorithm achieves approximately optimal performance:

Theorem 5 (Nemhauser [19]): Let G be a monotonic, submodular set function and $\mathcal{S}^* = \operatorname{argmax}_{\mathcal{S} \subseteq \mathcal{V}, |\mathcal{S}| \leq N} G(\mathcal{S})$.

Let \mathcal{S}^{gr} be the set retrieved by the greedy maximization Algorithm 1. We have the following guarantee for the performance of the greedy algorithm:

$$G(\mathcal{S}^{gr}) \geq (1 - (1 - \frac{1}{N})^N)G(\mathcal{S}^*) \geq (1 - \frac{1}{e})G(\mathcal{S}^*)$$

Moreover, no polynomial time algorithm can provide a better approximation guarantee unless $P=NP$ [39].

Combining the guarantees of Theorem 5, Lemma 1 and 2 we derive an approximation bound on the original problem (14):

Corollary 6:

$$(1 - \frac{1}{e}) \left[I(\tilde{f}_{\mathcal{S}^*}; \{\beta_m\}) - N \log \frac{\lambda \sigma_w^2 + 4\delta P(1+\delta)N^{\frac{3}{2}}}{\lambda \sigma_w^2 - \epsilon \lambda N^{\frac{3}{2}}} \right] \leq I(\hat{f}_{\mathcal{S}^{gr}}; \beta) \quad (23)$$

We hence apply Algorithm 1 to solve our optimization problem (18). The algorithm runs in time $O(|\mathcal{V}|N)$, linear in the size of the set \mathcal{V} and the number of selected elements N [40] such that it is easily implementable for problems of large size. However, as it turns out even more efficient variants of the algorithm have been introduced and studied. One such variant commonly referred to as the 'lazy greedy' algorithm was studied in [40] and it was shown to offer substantial running-time improvements in practice (with an unlikely worst-case theoretical performance upper bounded by that of the conventional greedy algorithm). Our numerical experiments described in Section V implement this more efficient variant to reduce running time.

3) *Improved Approximation Bounds*: While Theorem 5 guarantees an approximation bound of $(1 - \frac{1}{e}) \approx 63\%$ for the efficient greedy algorithm this guarantee is not tight. It is possible to derive a tighter data-dependent online bound on the gap between the cost of the greedy solution $G(\mathcal{S}^{gr})$ and that of the optimal solution $G(\mathcal{S}^*)$ [41]. Specifically, we have:

$$G(\mathcal{S}^{gr}) \geq G(\mathcal{S}^*) - \max_{B: |B| \leq N} \sum_{e \in B} (G(\mathcal{S}^{gr} \cup \{e\}) - G(\mathcal{S}^{gr})) \quad (24)$$

which takes $O(|\mathcal{V}| \log |\mathcal{V}|)$ evaluations of $G(\mathcal{S})$ to compute and sort. We use (24) to improve the distance from optimality bound in some of our numerical solutions in Section V.

F. Design Example: A Simple Ideal Setting

In the previous subsections we formulated the array design problem in a setting with constraints on the aperture \mathcal{A} and the number of sensors N and showed how a greedy algorithm (Algorithm 1) is guaranteed to efficiently find an approximate solution.

Here, we study a particular instance of that problem, where the Signal to Noise Ratio (SNR) is high, and the aperture is effectively unconstrained (the N sensors may be placed anywhere on the real line). Under these conditions a truncated $\frac{\lambda}{2}$ -spaced array is traditionally considered the design of choice in the conventional non-Bayesian setting (this is a truncated version of the infinite $\frac{\lambda}{2}$ -spaced design mentioned in Section II-A). We show next that the truncated $\frac{\lambda}{2}$ -spaced design naturally emerges as the approximately optimal solution as retrieved by our schemes in Bayesian settings where the a-priori β distribution satisfies some conditions. Specifically, we have the following result:

Theorem 7: Consider the high SNR regime $\frac{P}{\sigma_w^2} \rightarrow \infty$ and assume the prior from (7) takes a symmetric, monotonically decreasing form, i.e. $\sigma_m^2 = \sigma_{-m}^2$ and $\sigma_{m_1}^2 \geq \sigma_{m_2}^2$ whenever $0 \leq m_1 < m_2$. In addition, take \mathcal{V} as an arbitrarily dense set of sampling points on \mathbb{R} , and $\mathcal{M} = -M, \dots, M$ with $M \rightarrow \infty$.

We then have that a greedy solver on (18) will return a length N , $\frac{\lambda}{2}$ -spaced truncated uniform array centered around $x = 0$.

Proof: See Appendix C. \square

The last theorem studies one class of simple idealized problems where the greedy solution is reminiscent of generic non-Bayesian array designs. However, notice that our formalism is also useful in more challenging design problems such as when the aperture \mathcal{A} takes on arbitrary forms, and the effects of noise and application-tailored priors are considered. Furthermore, in Section IV we demonstrate that additional combinatorial constraints may be naturally incorporated into our formulation such that even more challenging problems become manageable.

IV. ARRAY DESIGN WITH MATROID CONSTRAINTS

In Section III we formalized the array design problem in a setting where we imposed constraints on the aperture and the number of sensors. Specifically, the constraints were $\mathcal{S} \subseteq \mathcal{V}$, $|\mathcal{S}| \leq N$. In many practical scenarios these may

be too simplistic to accurately represent real world design constraints. For example in applications where sensors are heavy and mounted on support beams we may want to restrict the number of sensors in specific sections of the aperture.

In this section we briefly review key elements from matroid theory which is a branch in combinatorics [42] and survey results from submodular optimization with matroid constraints guaranteeing the existence of efficient approximate solvers for this class of problems. We continue to show that these mathematical structures may be utilized to impose constraints of interest in array design enriching the set of problems our Bayesian formulation can describe and solve.

A. Submodular Optimization With Matroid Constraints

We begin by defining matroids and their corresponding independent sets [19]:

Definition 8: A finite **matroid** M is a pair $(\mathcal{V}, \mathcal{I})$ where \mathcal{V} is a ground set and \mathcal{I} is a collection of subsets of \mathcal{V} (the **independent sets**) that satisfies the following properties:

- 1) The empty set is independent: $\emptyset \in \mathcal{I}$
- 2) A subset of an independent set is independent:
 $X \subset Y, Y \in \mathcal{I} \Rightarrow X \in \mathcal{I}$
- 3) If X is an independent set and Y is a larger independent set, X can be augmented to a larger independent set by adding an element from $Y \setminus X$:

$$X, Y \in \mathcal{I}, |X| < |Y| \Rightarrow \exists e \in Y \setminus X \text{ s.t. } X \cup \{e\} \in \mathcal{I}$$

A matroid structure may be used to classify subsets of a ground set \mathcal{V} into permissible subsets which belong to \mathcal{I} and non permissible subsets which do not belong to \mathcal{I} . In the next subsection we show that using this formalism we can express interesting array design constraints.

From the theory of submodular optimization we have the following results for submodular optimization with matroid constraints [43], [44]. Let $M = (\mathcal{V}, \mathcal{I})$ be a matroid and $G(\mathcal{S})$ a monotonic, submodular set function. There exists an efficient approximate solver for the problem $\arg\max_{\mathcal{S} \in \mathcal{I}} G(\mathcal{S})$. Specifically, a greedy solver (maximizing the immediate marginal benefit at each step) taking the form

$$\mathcal{S}^{\text{gr}} \leftarrow \mathcal{S}^{\text{gr}} \cup \left\{ \arg\max_{e: e \notin \mathcal{S}^{\text{gr}}, \mathcal{S}^{\text{gr}} \cup \{e\} \in \mathcal{I}} [G(\mathcal{S}^{\text{gr}} \cup \{e\}) - G(\mathcal{S}^{\text{gr}})] \right\} \quad (25)$$

and stopping when no more elements e can be added is guaranteed to achieve a half-approximation bound:

$$G(\mathcal{S}^{\text{gr}}) \geq \frac{1}{2} \max_{\mathcal{S} \in \mathcal{I}} G(\mathcal{S}). \quad (26)$$

The constant factor may be tightened to $(1 - \frac{1}{e})$ by utilizing specialized randomized algorithms [44].

B. Combinatorial Constraints in Array Design

Here we invoke one specific well known matroid structure and show its application in expressing useful array design constraints. Let \mathcal{V} be a ground set of grid points where sensors are allowed to be placed as before. Let $\mathcal{V}_1, \dots, \mathcal{V}_K$ be a partition of the set \mathcal{V} , i.e. $\bigcup_k \mathcal{V}_k = \mathcal{V}$, $\mathcal{V}_i \cap \mathcal{V}_j = \emptyset$, $\forall i \neq j$, and let N, n_1, \dots, n_K be a set of integers.

We define the (cardinality constrained) partition matroid [44] $M = (\mathcal{V}, \mathcal{I})$ with the following definition for the collection of independent sets: a subset $\mathcal{S} \subseteq \mathcal{V}$ is an independent subset $\mathcal{S} \in \mathcal{I}$ if it holds $|\mathcal{S} \cap \mathcal{V}_i| \leq n_i$, $\forall i$.

In the context of array design the partition matroid may be useful in expressing practical constraints over sensor placement configurations. For example if the subsets \mathcal{V}_i represent closed line sections, e.g. a physical partitioning of the aperture into zones, and \mathcal{I} represents the collection of all permissible designs then the structure of the matroid limits the number of sensors that may be placed in the i th zone to n_i which may be an important engineering constraint coupled with some specific application. We solve:

$$\mathcal{S}^* = \arg\max_{\mathcal{S} \in \mathcal{I}} I(\hat{\mathbf{f}}_{\mathcal{S}}; \boldsymbol{\beta}) \quad (27)$$

Applying the results from the previous subsection we immediately have an efficient approximate solver for the array design problem coupled with a partition matroid constraint. In Section V we detail such a design for one numerical example.

V. NUMERICAL EXPERIMENTS

In this section we perform numerical experiments validating our theoretical results and exemplifying them. We showcase an array design with cardinality and aperture constraints as prescribed in Section III and design arrays adhering to matroid constraints as prescribed in Section IV.

Our initial setting is as follow. We fix $\lambda = 1$ throughout as the wavelength only serves to scale the x axis. The aperture is set as $\mathcal{A} = \{x | -3.5 \leq x \leq 3.5\}$ and the selection set \mathcal{V} is chosen as a uniform grid of 113 positions from \mathcal{A} spaced $\delta = 0.0625$ apart. We set out to design an array consisting of $N = 11$ sensor locations. The prior for $\{\beta_m\}$ is set as per (9) with $r = 1$ and normalized to sum to $P = 1$. For the simulations we consider the truncated vector $\boldsymbol{\beta}$ formed when restricting the set of m coefficients to 901 consecutive elements centered around the origin, i.e., we set $\mathcal{M} = \{-450, \dots, +450\}$. For the preliminary design we implement the lazy greedy algorithm of Section III and plot the results in the left column of Fig. 2 as a function of the SNR which we define here as $\frac{P}{N\sigma_p^2}$. Blue markers denote the full selection set \mathcal{V} and red markers delineate the active \mathcal{S} selected by the algorithm.

In the high SNR regime (SNR = 10dB or higher values) the resulting design is a truncated $\frac{1}{2}$ uniform array as predicted according to Theorem 7. As the SNR decreases the reliability of the measurements deteriorates and the algorithm prefers locating samplers right next to each other on expense of widening the array as this serves to average out the noise.

The performance in terms of mutual information $I(\hat{\mathbf{f}}_{\mathcal{S}}; \boldsymbol{\beta})$ for the selected locations \mathcal{S} appears in the title of the plots (in natural units). Notice for example that for the 5dB SNR design the achieved mutual information is 12.54. Using Theorem 5 we have that the optimal design cannot achieve mutual information better than $\frac{1}{1 - \frac{1}{e}} 12.54 = 19.83$. This bound can be improved using the improved bounding method briefly

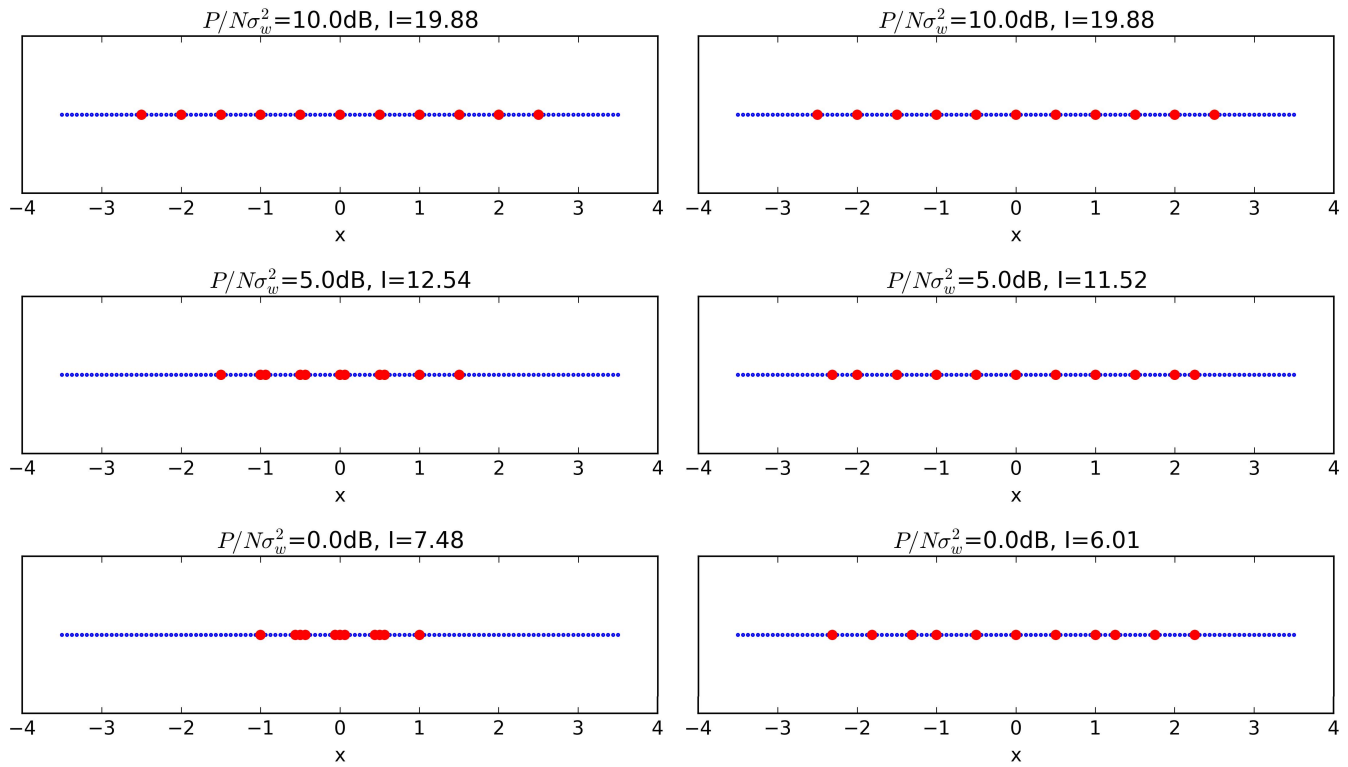


Fig. 2. (left) Array designs with an aperture constraint and various SNR levels. (right) Array designs with combinatorial placement constraints and various SNR levels.

described following Theorem 5 to show that the optimal performance is not greater than 17.45.

The truncation level dictated by our choice of \mathcal{M} translates to $\epsilon \cong 1e-4$ and the truncation bounds from Lemma 1 read (for the lower, extreme SNR case): $-0.45 \leq I(\hat{f}_{\mathcal{S}}; \{\beta_m\}) - I(\hat{f}_{\mathcal{S}}; \beta) \leq 0.47$. We find empirically that these bounds are extremely loose and \mathcal{M} can be shrunk considerably without substantially compromising accuracy. To achieve a similar upper bound on $I(\hat{f}_{\mathcal{S}^*}; \beta) - I(\hat{f}_{\mathcal{S}_d^*}; \beta)$ as per Lemma 2 a discretization level of $\delta \cong 2e-4$ is needed. However, we empirically find that our choice of $\delta = 0.0625$ is accurate enough as further refining the grid does not significantly change the design. Our lemmas prove to be extremely pessimistic as is expected given that the proofs take into account worst-case scenarios.

The array geometries above, derived according to the formulations of Section III, are designed to optimize the quality of inference between the measurements and the scene expansion coefficients β . Many sensing applications of interest specifically involve imaging the scene, that is reconstructing $\beta(\psi)$ from the measurements. Our next experiment was designed to empirically evaluate the Mean Square Error (MSE) performance in scene reconstruction from measurements collected using the prescribed designs. First, we designed five array geometries as described above, optimized for several target SNR levels {30dB, 12dB, 10dB, 5dB, 0dB}. We set up a Monte-Carlo experiment where 1000 scenes were randomly drawn from the distribution of Section II-B.2. For each scene, noisy measurements were collected by each of the five

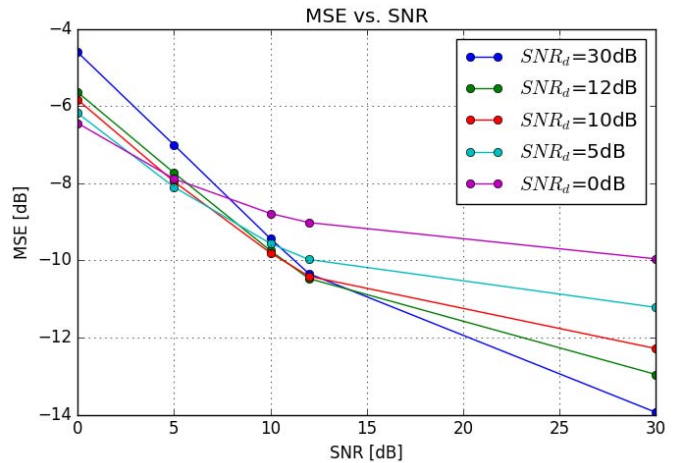


Fig. 3. Reconstruction MSE for various arrays designed for fixed SNR levels, with actual SNR levels swept.

optimized arrays. The measurements were repeated with five different synthetic noise levels corresponding to the five target SNR levels.

We repeatedly performed maximum likelihood estimation [45] of the expansion coefficients β , and synthesized an estimated scene $\hat{\beta}(\psi)$ according to (4). The MSE discrepancy between $\hat{\beta}(\psi)$ and the true scene is depicted in Fig. 3. It is evident that the quality of inference criterion is indicative of MSE performance, as each of the five geometries yielded the best MSE performance at its specified target SNR level.

Next we solve a corresponding set of design problems with matroid constraints installed to limit the number of sensors in given aperture segments. Specifically, we use the (cardinality constrained) partition matroid from Section IV-B with $N = 11$, $n_i = 1, \forall i$ and \mathcal{V}_i spanning consecutive line segments of length 0.5: $\mathcal{V}_i = [-0.25 + 0.5 \cdot i, +0.25 + 0.5 \cdot i)$. The matroid constraints limits the proximity between sensor elements, which may be useful in real world use cases. We plot the results for the matroid constrained designs in the right column of Fig. 2. Notice that while the theoretical guarantees pertaining to the greedy matroid optimization scheme of Section IV is 0.5 compared to $1 - \frac{1}{e}$ with the cardinality constraints of Section III, the actual performance achieved in the constrained design instances is not far from those achieved with the simple cardinality constraints.

VI. CONCLUDING REMARKS

We introduced a novel framework for designing sensor arrays. Our setting is Bayesian in the sense that our model incorporates the notion of prior beliefs about the scene of interest. Our goal is designing arrays that are adapted to the scene and perform efficient inference. We showed that this NP-hard combinatorial selection problem may be efficiently approximated by porting results from the theory of submodular set function optimization.

Initially, we showed how to apply our formalism in design problems with straightforward cardinality and aperture constraints. Later we showed that more challenging combinatorial constraints may be enforced by utilizing results and concepts from the field of matroids and submodular optimization with matroid constraints.

Our formulation connecting the array design problem and submodularity may further be extended to reap additional benefits, namely tackling problems such as robust array design or adaptive design that evolve as the scene is learned. We leave the treatment of some of these and other problems to separate study [46].

APPENDIX A PROOF OF LEMMA 1

For simplicity of notation, we suppress the subscript \mathcal{S} throughout. Begin by expanding the mutual information expressions:

$$\begin{aligned} I(\tilde{\mathbf{f}}; \{\beta_m\}) &= H(\tilde{\mathbf{f}}) - H(\tilde{\mathbf{f}} | \{\beta_m\}) \\ I(\hat{\mathbf{f}}; \boldsymbol{\beta}) &= H(\hat{\mathbf{f}}) - H(\hat{\mathbf{f}} | \boldsymbol{\beta}) \end{aligned} \quad (28)$$

Examining (12) and (15) we have $H(\tilde{\mathbf{f}} | \{\beta_m\}) = H(\hat{\mathbf{f}} | \boldsymbol{\beta}) = H(\mathbf{w})$ and so:

$$I(\tilde{\mathbf{f}}; \{\beta_m\}) - I(\hat{\mathbf{f}}; \boldsymbol{\beta}) = H(\tilde{\mathbf{f}}) - H(\hat{\mathbf{f}}) \quad (29)$$

Next, notice that $\tilde{\mathbf{f}}$ and $\hat{\mathbf{f}}$ are both circular, complex, Gaussian random N -length vectors, such that their entropies are given according to [27]:

$$\begin{aligned} H(\tilde{\mathbf{f}}) &= \log((\pi e)^N \det(\tilde{\boldsymbol{\Sigma}})) \quad \tilde{\Sigma}_{ij} = \mathbb{E}[\tilde{\mathbf{f}}_i \tilde{\mathbf{f}}_j^*] \\ H(\hat{\mathbf{f}}) &= \log((\pi e)^N \det(\hat{\boldsymbol{\Sigma}})) \quad \hat{\Sigma}_{ij} = \mathbb{E}[\hat{\mathbf{f}}_i \hat{\mathbf{f}}_j^*] \end{aligned} \quad (30)$$

Further expand using the independence between \mathbf{w} and β_m and $\mathbb{E}[\beta_m \beta_{m'}^*] = \delta_{mm'} \sigma_m^2$:

$$\begin{aligned} \tilde{\Sigma}_{ij} &= \mathbb{E}[(\sum_m K_{im} \beta_m + w_i)(\sum_{m'} K_{jm'}^* \beta_{m'}^* + w_j^*)] \\ &= [\boldsymbol{\Sigma}_{\mathbf{w}\mathbf{w}}]_{ij} + \sum_m K_{im} K_{jm}^* \sigma_m^2 \end{aligned} \quad (31)$$

$$\begin{aligned} \hat{\Sigma}_{ij} &= \mathbb{E}[(\sum_{m \in \mathcal{M}} K_{im} \beta_m + w_i)(\sum_{m' \in \mathcal{M}} K_{jm'}^* \beta_{m'}^* + w_j^*)] \\ &= [\boldsymbol{\Sigma}_{\mathbf{w}\mathbf{w}}]_{ij} + \sum_{m \in \mathcal{M}} K_{im} K_{jm}^* \sigma_m^2 \end{aligned} \quad (32)$$

Comparing (31) and (32) and using:

$$\left| \sum_{m \notin \mathcal{M}} K_{im} K_{jm}^* \sigma_m^2 \right| \leq \sum_{m \notin \mathcal{M}} |K_{im} K_{jm}^* \sigma_m^2| \leq \sum_{m \notin \mathcal{M}} \sigma_m^2 = \epsilon \quad (33)$$

we have $|\tilde{\Sigma}_{ij} - \hat{\Sigma}_{ij}| \leq \epsilon$ such that we may write:

$$\hat{\boldsymbol{\Sigma}} = \tilde{\boldsymbol{\Sigma}} + \epsilon \mathbf{X} \quad (34)$$

for some $N \times N$ matrix \mathbf{X} satisfying $|X_{ij}| \leq 1$. We use (34) to bound the determinants. $\tilde{\boldsymbol{\Sigma}}$ is positive-definite and invertible such that we can write:

$$\det(\hat{\boldsymbol{\Sigma}}) = \det(\tilde{\boldsymbol{\Sigma}} + \epsilon \mathbf{X}) = \det(\tilde{\boldsymbol{\Sigma}}) \det(\mathbf{I}_N + \epsilon \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{X}) \quad (35)$$

Substituting (35) in (30) we have:

$$H(\hat{\mathbf{f}}) - H(\tilde{\mathbf{f}}) = \log(\det(\mathbf{I}_N + \epsilon \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{X})) = \log(\det(\tilde{\mathbf{X}})) \quad (36)$$

with $\tilde{\mathbf{X}} \equiv \mathbf{I}_N + \epsilon \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{X}$. We turn next to bounding the term $\log(\det(\tilde{\mathbf{X}}))$. First notice:

$$\begin{aligned} \left| [\epsilon \tilde{\boldsymbol{\Sigma}}^{-1} \mathbf{X}]_{ij} \right| &= \epsilon \left| \sum_m \tilde{\Sigma}_{im}^{-1} X_{mj} \right| \leq \epsilon \sum_m \left| \tilde{\Sigma}_{im}^{-1} X_{mj} \right| \\ &\leq \epsilon \sum_m \left| \tilde{\Sigma}_{im}^{-1} \right| \leq \epsilon \|\tilde{\boldsymbol{\Sigma}}^{-1}\|_\infty \leq \epsilon \sqrt{N} \|\tilde{\boldsymbol{\Sigma}}^{-1}\|_2 \\ &= \epsilon \sqrt{N} \sigma_{\max}(\tilde{\boldsymbol{\Sigma}}^{-1}) = \epsilon \sqrt{N} \frac{1}{\sigma_{\min}(\tilde{\boldsymbol{\Sigma}})} \leq \frac{\epsilon \sqrt{N}}{\sigma_w^2} \end{aligned} \quad (37)$$

Where we have used the matrix norm equivalence $\|\mathbf{A}\|_\infty \leq \sqrt{N} \|\mathbf{A}\|_2$ (for $N \times N$ matrices) and $\sigma_{\max}(\cdot)$ ($\sigma_{\min}(\cdot)$) is the maximal (minimal) singular value such that $\sigma_{\min}(\tilde{\boldsymbol{\Sigma}}) \geq \sigma_w^2$. Thus we have that $\tilde{\mathbf{X}}$ has diagonal elements centered around 1: $|\tilde{X}_{ii} - 1| \leq \frac{\epsilon \sqrt{N}}{\sigma_w^2}$ and the row-sums over non-diagonal entries satisfy $\sum_{m \neq i} |\tilde{X}_{im}| \leq \frac{\epsilon \sqrt{N(N-1)}}{\sigma_w^2}$.

Applying the Gershgorin circle theorem we have for the eigenvalues of $\tilde{\mathbf{X}}$:

$$1 - \frac{\epsilon \sqrt{N} N}{\sigma_w^2} \leq |\lambda_i| \leq 1 + \frac{\epsilon \sqrt{N} N}{\sigma_w^2}$$

$\det(\tilde{\mathbf{X}})$ is a positive real number as the quotient of the determinants of two positive definite matrices $\det(\tilde{\mathbf{X}}) = \frac{\det(\hat{\boldsymbol{\Sigma}})}{\det(\tilde{\boldsymbol{\Sigma}})}$

such that we can write $\det(\tilde{\mathbf{X}}) = \prod_i \lambda_i = \prod_i |\lambda_i|$ and consequently:

$$N \log\left(1 - \frac{\epsilon N^{\frac{3}{2}}}{\sigma_w^2}\right) \leq \log(\det(\tilde{\mathbf{X}})) \leq N \log\left(1 + \frac{\epsilon N^{\frac{3}{2}}}{\sigma_w^2}\right) \quad (38)$$

This finally leads to:

$$\begin{aligned} -N \log\left(1 + \frac{\epsilon N^{\frac{3}{2}}}{\sigma_w^2}\right) &\leq I(\tilde{\mathbf{f}}; \{\beta_m\}) - I(\hat{\mathbf{f}}; \boldsymbol{\beta}) \\ &\leq -N \log\left(1 - \frac{\epsilon N^{\frac{3}{2}}}{\sigma_w^2}\right) \end{aligned} \quad (39)$$

APPENDIX B PROOF OF LEMMA 2

The left inequality is trivial. We have $I(\hat{\mathbf{f}}_{\mathcal{S}^*}; \boldsymbol{\beta}) \leq I(\hat{\mathbf{f}}_{\mathcal{S}}; \boldsymbol{\beta})$ as the second optimization is over a larger set.

To prove the right inequality we show that for every $\mathcal{S} \subseteq \mathcal{A}$ there is $\mathcal{S}_d \subseteq \mathcal{V}$ such that

$$I(\hat{\mathbf{f}}_{\mathcal{S}}; \boldsymbol{\beta}) \leq I(\hat{\mathbf{f}}_{\mathcal{S}_d}; \boldsymbol{\beta}) + N \log\left(1 + \frac{4\delta P(1 + \frac{\delta}{\lambda})N^{\frac{3}{2}}}{\lambda\sigma_w^2}\right) \quad (40)$$

we will show this for $|\mathcal{S}|=N$ but the proof for other cardinalities is identical.

With distance δ between adjacent elements of \mathcal{V} , for every $\mathcal{S} = \{x_1, \dots, x_N\} \subset \mathcal{A}$ there is a set $\mathcal{S}_d = \{x_1^d, \dots, x_N^d\}$ such that $\mathcal{S}_d \subseteq \mathcal{V}$ and $|x_i - x_i^d| \leq \frac{\delta}{2}$ for all i . We have, similarly to Appendix A:

$$I(\hat{\mathbf{f}}_{\mathcal{S}}; \boldsymbol{\beta}) - I(\hat{\mathbf{f}}_{\mathcal{S}_d}; \boldsymbol{\beta}) = H(\hat{\mathbf{f}}_{\mathcal{S}}) - H(\hat{\mathbf{f}}_{\mathcal{S}_d}) \quad (41)$$

Using the model (15):

$$\begin{aligned} H(\hat{\mathbf{f}}_{\mathcal{S}}) &= \log((\pi e)^N \det(\hat{\boldsymbol{\Sigma}}^{\mathcal{S}})) \\ H(\hat{\mathbf{f}}_{\mathcal{S}_d}) &= \log((\pi e)^N \det(\hat{\boldsymbol{\Sigma}}^{\mathcal{S}_d})) \end{aligned} \quad (42)$$

where:

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}^{\mathcal{S}} &\equiv \mathbb{E}[\hat{\mathbf{f}}_{\mathcal{S}} \hat{\mathbf{f}}_{\mathcal{S}}^\dagger] = \mathbf{K}_{\mathcal{S}} \boldsymbol{\Sigma}_{\beta\beta} \mathbf{K}_{\mathcal{S}}^\dagger + \boldsymbol{\Sigma}_{ww} \\ \hat{\boldsymbol{\Sigma}}^{\mathcal{S}_d} &\equiv \mathbb{E}[\hat{\mathbf{f}}_{\mathcal{S}_d} \hat{\mathbf{f}}_{\mathcal{S}_d}^\dagger] = \mathbf{K}_{\mathcal{S}_d} \boldsymbol{\Sigma}_{\beta\beta} \mathbf{K}_{\mathcal{S}_d}^\dagger + \boldsymbol{\Sigma}_{ww} \\ \boldsymbol{\Sigma}_{\beta\beta} &\equiv \mathbb{E}[\boldsymbol{\beta} \boldsymbol{\beta}^\dagger] \end{aligned} \quad (43)$$

and $H(\hat{\mathbf{f}}_{\mathcal{S}}) - H(\hat{\mathbf{f}}_{\mathcal{S}_d}) = \log(\det(\hat{\boldsymbol{\Sigma}}^{\mathcal{S}})) - \log(\det(\hat{\boldsymbol{\Sigma}}^{\mathcal{S}_d}))$.

Both $\mathbf{K}_{\mathcal{S}}$ and $\mathbf{K}_{\mathcal{S}_d}$ are size $N \times |\mathcal{M}|$ matrices defined as per the definition in (11), such that:

$$\begin{aligned} |[\mathbf{K}_{\mathcal{S}}]_{nm} - [\mathbf{K}_{\mathcal{S}_d}]_{nm}| &= |\text{sinc}(m + \frac{2}{\lambda}x_n) - \text{sinc}(m + \frac{2}{\lambda}x_n^d)| \\ &\leq 2\frac{2}{\lambda}|x_n - x_n^d| \leq \frac{2}{\lambda}\delta \end{aligned} \quad (44)$$

where for the first inequality we have used the fact that $\text{sinc}(\cdot)$ is Lipschitz with constant smaller than 2. We can thus define $\boldsymbol{\Delta} \equiv \mathbf{K}_{\mathcal{S}} - \mathbf{K}_{\mathcal{S}_d}$ and we have $|\Delta_{nm}| \leq \frac{2}{\lambda}\delta$. Substitution in (43) yields:

$$\hat{\boldsymbol{\Sigma}}^{\mathcal{S}} = \hat{\boldsymbol{\Sigma}}^{\mathcal{S}_d} + \boldsymbol{\Delta} \boldsymbol{\Sigma}_{\beta\beta} \boldsymbol{\Delta}^\dagger + \boldsymbol{\Delta} \boldsymbol{\Sigma}_{\beta\beta} \mathbf{K}_{\mathcal{S}_d}^\dagger + \mathbf{K}_{\mathcal{S}_d} \boldsymbol{\Sigma}_{\beta\beta} \boldsymbol{\Delta}^\dagger \quad (45)$$

We bound the perturbation terms by noticing:

$$\begin{aligned} |[\boldsymbol{\Delta} \boldsymbol{\Sigma}_{\beta\beta} \boldsymbol{\Delta}^\dagger]_{ij}| &= \left| \sum_m \Delta_{im} [\boldsymbol{\Sigma}_{\beta\beta}]_{mm} \Delta_{jm} \right| \\ &\leq \left(\frac{2}{\lambda}\delta\right)^2 \sum_m [\boldsymbol{\Sigma}_{\beta\beta}]_{mm} \leq \frac{4}{\lambda^2} \delta^2 P \\ |[\boldsymbol{\Delta} \boldsymbol{\Sigma}_{\beta\beta} \mathbf{K}_{\mathcal{S}_d}^\dagger]_{ij}| &= \left| \sum_m \Delta_{im} [\boldsymbol{\Sigma}_{\beta\beta}]_{mm} [\mathbf{K}_{\mathcal{S}_d}]_{jm} \right| \\ &\leq \frac{2}{\lambda} \delta \cdot 1 \cdot \sum_m [\boldsymbol{\Sigma}_{\beta\beta}]_{mm} \leq \frac{2}{\lambda} \delta P \end{aligned} \quad (46)$$

and overall we have:

$$|\hat{\boldsymbol{\Sigma}}_{ij}^{\mathcal{S}} - \hat{\boldsymbol{\Sigma}}_{ij}^{\mathcal{S}_d}| \leq \frac{4}{\lambda} \delta P + \frac{4}{\lambda^2} \delta^2 P = \frac{4}{\lambda} \delta P \left(1 + \frac{\delta}{\lambda}\right) \quad (47)$$

define: $\epsilon' \equiv \frac{4}{\lambda} \delta P \left(1 + \frac{\delta}{\lambda}\right)$ and we have

$$\hat{\boldsymbol{\Sigma}}^{\mathcal{S}} = \hat{\boldsymbol{\Sigma}}^{\mathcal{S}_d} + \epsilon' \mathbf{X} \quad (48)$$

with $N \times N$ matrix \mathbf{X} satisfying $|X_{ij}| \leq 1$ which is akin to (34). We thus port the results from Appendix A here (we only need the lower bound):

$$-N \log\left(1 + \frac{\epsilon' N^{\frac{3}{2}}}{\sigma_w^2}\right) \leq I(\hat{\mathbf{f}}_{\mathcal{S}_d}; \boldsymbol{\beta}) - I(\hat{\mathbf{f}}_{\mathcal{S}}; \boldsymbol{\beta}) \quad (49)$$

which, upon substitution of ϵ' is equivalent to (40).

APPENDIX C PROOF OF THEOREM 7

The greedy algorithm sequentially selects elements according to the rule $x^* = \arg\max_{x \in \mathcal{V} \setminus \mathcal{S}} I(\hat{\mathbf{f}}_{\mathcal{S} \cup \{x\}}; \boldsymbol{\beta})$ where \mathcal{S} is the set of elements selected so far. We recursively show that the added elements can be selected on a $\frac{\lambda}{2}$ -spaced grid centered around $x = 0$. Expanding the mutual information as in Appendix A we have:

$$\arg\max_{x \in \mathcal{V} \setminus \mathcal{S}} I(\hat{\mathbf{f}}_{\mathcal{S} \cup \{x\}}; \boldsymbol{\beta}) = \arg\max_{x \in \mathcal{V} \setminus \mathcal{S}} H(\hat{\mathbf{f}}_{\mathcal{S} \cup \{x\}}) \quad (50)$$

We begin by showing that the first selected element is $x_1 = 0$. Indeed, using the results from Appendix A:

$$\arg\max_{x \in \mathcal{V}} H(\hat{\mathbf{f}}_{\{x\}}) = \arg\max_{x \in \mathcal{V}} \det(\hat{\boldsymbol{\Sigma}}_{11}) = \arg\max_{x \in \mathcal{V}} \hat{\boldsymbol{\Sigma}}_{11} \quad (51)$$

where again using Appendix A and under the assumptions of the theorem (high SNR):

$$\begin{aligned} \hat{\boldsymbol{\Sigma}}_{11} &= \sum_m K_{1m} K_{1m}^* \sigma_m^2 = \sum_m \text{sinc}^2\left(m + \frac{2}{\lambda}x\right) \sigma_m^2 \\ &\leq \sigma_0^2 \sum_m \text{sinc}^2\left(m + \frac{2}{\lambda}x\right) = \sigma_0^2 \end{aligned} \quad (52)$$

where we used $\sigma_m^2 \leq \sigma_0^2$, $\forall m \neq 0$ and the identity:

$$\sum_m \text{sinc}(m+a) \text{sinc}(m+b) = \text{sinc}(b-a) \quad (53)$$

It easy to see that in (52) equality is achieved for the choice $x_1 = 0$ which is the claim.

Next, assume that the greedy algorithm has already picked a set $\mathcal{S} = \{x_1, \dots, x_{|\mathcal{S}|}\}$ of adjacent elements on a $\frac{\lambda}{2}$ -spaced

grid centered around $x = 0$ and show that the next element to be added is an adjacent location on the same $\frac{\lambda}{2}$ -spaced grid. We have using $H(x, y) = H(x) + H(y|x)$:

$$\operatorname{argmax}_{x \in \mathcal{V} \setminus \mathcal{S}} H(\hat{\mathbf{f}}_{\mathcal{S} \cup \{x\}}) = \operatorname{argmax}_{x \in \mathcal{V} \setminus \mathcal{S}} H(\hat{\mathbf{f}}_{\{x\}} | \hat{\mathbf{f}}_{\mathcal{S}}) = \operatorname{argmax}_{x \in \mathcal{V} \setminus \mathcal{S}} \sigma_{x|\mathcal{S}}^2 \quad (54)$$

where $\sigma_{x|\mathcal{S}}^2$ is the conditional variance of the Gaussian observation collected at x given the Gaussian observations made at the set \mathcal{S} :

$$\sigma_{x|\mathcal{S}}^2 = \sigma_x^2 - \boldsymbol{\Sigma}_{x\mathcal{S}} \boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}^{-1} \boldsymbol{\Sigma}_{\mathcal{S}x}^\dagger \quad (55)$$

with the usual definitions for the covariance matrices (as in Appendix A):

$$\begin{aligned} [\boldsymbol{\Sigma}_{x\mathcal{S}}]_{li} &= \sum_m \operatorname{sinc}(m + \frac{2}{\lambda}x) \operatorname{sinc}(m + \frac{2}{\lambda}x_i) \sigma_m^2 \\ &= \operatorname{sinc}(m(i) + \frac{2}{\lambda}x) \sigma_{m(i)}^2 \\ [\boldsymbol{\Sigma}_{\mathcal{S}\mathcal{S}}]_{ij} &= \sum_m \operatorname{sinc}(m + \frac{2}{\lambda}x_i) \operatorname{sinc}(m + \frac{2}{\lambda}x_j) \sigma_m^2 = \delta_{ij} \sigma_{m(i)}^2 \end{aligned} \quad (56)$$

where in the last equations almost all $\operatorname{sinc}(\cdot)$ functions nulled out as the x_i 's are situated on a $\frac{\lambda}{2}$ grid, and we have defined $m(i) \equiv -\frac{2}{\lambda}x_i$, such that $I \equiv \{m(i)\}$ is a set of consecutive integers. Additionally, we have:

$$\sigma_x^2 = \sum_m \operatorname{sinc}^2(m + \frac{2}{\lambda}x) \sigma_m^2 \quad (57)$$

Substituting back into (55) we have:

$$\begin{aligned} \sigma_{x|\mathcal{S}}^2 &= \sum_m \operatorname{sinc}^2(m + \frac{2}{\lambda}x) \sigma_m^2 - \sum_{i \in I} \operatorname{sinc}^2(m(i) + \frac{2}{\lambda}x) \sigma_{m(i)}^2 \\ &= \sum_{m \notin I} \operatorname{sinc}^2(m + \frac{2}{\lambda}x) \sigma_m^2 \end{aligned} \quad (58)$$

and this is similar to the optimization over the selection of the first location in (52) with the optimum achieved by selecting x such that $\frac{2}{\lambda}x = m$ for the first $m \notin I$ which is the next adjacent location on the $\frac{\lambda}{2}$ grid which completes the proof.

REFERENCES

- [1] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*. Hoboken, NJ, USA: Wiley, 2004.
- [2] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul. 1996.
- [3] R. L. Haupt, "Thinned arrays using genetic algorithms," *IEEE Trans. Antennas Propag.*, vol. 42, no. 7, pp. 993–999, Jul. 1994.
- [4] M. M. Khodier and C. G. Christodoulou, "Linear array geometry synthesis with minimum sidelobe level and null control using particle swarm optimization," *IEEE Trans. Antennas Propag.*, vol. 53, no. 8, pp. 2674–2679, Aug. 2005.
- [5] M. I. Skolnik, G. Nemhauser, and J. Sherman, III, "Dynamic programming applied to unequally spaced arrays," *IEEE Trans. Antennas Propag.*, vol. AP-12, no. 1, pp. 35–43, Jan. 1964.
- [6] R. L. Haupt and D. H. Werner, *Genetic Algorithms in Electromagnetics*. Hoboken, NJ, USA: Wiley, 2007.
- [7] B. P. Kumar and G. R. Branner, "Design of unequally spaced arrays for performance improvement," *IEEE Trans. Antennas Propag.*, vol. 47, no. 3, pp. 511–523, Mar. 1999.
- [8] G. Oliveri, M. Carlin, and A. Massa, "Complex-weight sparse linear array synthesis by Bayesian compressive sampling," *IEEE Trans. Antennas Propag.*, vol. 60, no. 5, pp. 2309–2326, May 2012.
- [9] H. Gazzah and S. Marcos, "Cramér-Rao bounds for antenna array design," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 336–345, Jan. 2006.
- [10] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 573–586, Feb. 2011.
- [11] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, no. 10, pp. 1783–1795, Oct. 1990.
- [12] M. Wang and A. Nehorai, "Coarrays, MUSIC, and the Cramér-Rao bound," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 933–946, Feb. 2017.
- [13] A. Koochakzadeh and P. Pal, "Cramér-Rao bounds for underdetermined source localization," *IEEE Signal Process. Lett.*, vol. 23, no. 7, pp. 919–923, Jul. 2016.
- [14] D. Romero, R. López-Valcarce, and G. Leus, "Compression limits for random vectors with linearly parameterized second-order statistics," *IEEE Trans. Inf. Theory*, vol. 61, no. 3, pp. 1410–1425, Mar. 2015.
- [15] D. D. Ariananda and G. Leus, "Compressive wideband power spectrum estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4775–4789, Sep. 2012.
- [16] A. Koochakzadeh and P. Pal, "Performance of uniform and sparse non-uniform samplers in presence of modeling errors: A Cramér-Rao bound based study," *IEEE Trans. Signal Process.*, vol. 65, no. 6, pp. 1607–1621, Mar. 2017.
- [17] J. D. Krieger, Y. Kochman, and G. W. Wornell, "Design and analysis of multi-coset arrays," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 3781–3785.
- [18] J. D. Krieger, Y. Kochman, and G. W. Wornell, "Multi-coset sparse imaging arrays," *IEEE Trans. Antennas Propag.*, vol. 62, no. 4, pp. 1701–1715, Apr. 2014.
- [19] G. L. Nemhauser, L. A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximizing submodular set functions—I," *Math. Program.*, vol. 14, no. 1, pp. 265–294, 1978.
- [20] S. Fujishige, *Submodular Functions and Optimization*, vol. 58. Amsterdam, The Netherlands: Elsevier, 2005.
- [21] N. Buchbinder, M. Feldman, J. S. Naor, and R. Schwartz, "A tight linear time (1/2)-approximation for unconstrained submodular maximization," *SIAM J. Comput.*, vol. 44, no. 5, pp. 1384–1402, 2015.
- [22] J. Vondrák, "Symmetry and approximability of submodular maximization problems," *SIAM J. Comput.*, vol. 42, no. 1, pp. 265–304, 2013.
- [23] B. Mirzasoleiman, A. Karbasi, R. Sarkar, and A. Krause, "Distributed submodular maximization: Identifying representative elements in massive data," in *Proc. Adv. Neural Inf. Process. Syst.*, 2013, pp. 2049–2057.
- [24] A. Krause, A. Singh, and C. Guestrin, "Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies," *J. Mach. Learn. Res.*, vol. 9, no. 2, pp. 235–284, 2008.
- [25] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," in *Proc. 49th IEEE Conf. Decis. Control (CDC)*, Dec. 2010, pp. 2572–2577.
- [26] A. Krause and C. Guestrin, "Submodularity and its applications in optimized information gathering," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 4, 2011, Art. no. 32.
- [27] F. D. Neeser and J. L. Massey, "Proper complex random processes with applications to information theory," *IEEE Trans. Inf. Theory*, vol. 39, no. 4, pp. 1293–1302, Jul. 1993.
- [28] J. Yen, "On nonuniform sampling of bandwidth-limited signals," *IRE Trans. Circuit Theory*, vol. 3, no. 4, pp. 251–257, Dec. 1956.
- [29] D. J. Wingham, "The reconstruction of a band-limited function and its Fourier transform from a finite number of samples at arbitrary locations by singular value decomposition," *IEEE Trans. Signal Process.*, vol. 40, no. 3, pp. 559–570, Mar. 1992.
- [30] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2015.
- [31] M. K. Mihçak, I. Kozintsev, K. Ramchandran, and P. Moulin, "Low-complexity image denoising based on statistical modeling of wavelet coefficients," *IEEE Signal Process. Lett.*, vol. 6, no. 12, pp. 300–303, Dec. 1999.
- [32] A. Levin, R. Fergus, F. Durand, and W. T. Freeman, "Image and depth from a conventional camera with a coded aperture," *ACM Trans. Graph.*, vol. 26, no. 3, p. 70, 2007.
- [33] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2009, pp. 1964–1971.
- [34] A. Kulesza and B. Taskar, "Determinantal point processes for machine learning," *Found. Trends Mach. Learn.*, vol. 5, nos. 2–3, pp. 123–286, 2012.

- [35] G. Friesecke, "Lecture notes in Fourier analysis [MA4064]," Technische Univ. München, Munich, Germany, Tech. Rep., 2013.
- [36] K. Chaloner and I. Verdine, "Bayesian experimental design: A review," *Stat. Sci.*, vol. 10, no. 3, pp. 273–304, 1995.
- [37] J. M. Bernardo, "Expected information as expected utility," *Ann. Statist.*, vol. 7, no. 3, pp. 686–690, 1979.
- [38] A. Krause and C. Guestrin, "Near-optimal nonmyopic value of information in graphical models," in *Proc. 21st Conf. Uncertainty Artif. Intell.*, 2012, pp. 324–331.
- [39] U. Feige, "A threshold of $\ln n$ for approximating set cover," *J. ACM*, vol. 45, no. 4, pp. 634–652, 1998.
- [40] M. Minoux, "Accelerated greedy algorithms for maximizing submodular set functions," in *Optimization Techniques*. Berlin, Germany: Springer, 1978, pp. 234–243.
- [41] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2007, pp. 420–429.
- [42] J. G. Oxley, *Matroid Theory*, vol. 3. New York, NY, USA: Oxford Univ. Press, 2006.
- [43] A. Krause and D. Golovin, "Submodular function maximization," in *Tractability: Practical Approaches to Hard Problems*. Cambridge, U.K.: Cambridge Univ. Press, 2014, pp. 71–104.
- [44] G. Calinescu, C. Chekuri, M. Pál, and J. Vondrák, "Maximizing a monotone submodular function subject to a matroid constraint," *SIAM J. Comput.*, vol. 40, no. 6, pp. 1740–1766, 2011.
- [45] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5101–5125, Nov. 2011.
- [46] G. Shulkind, S. Jegelka, and G. W. Wornell, "Multiple wavelength sensing array design," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2017, pp. 3424–3428.

Gal Shulkind received a B.A. in Physics and a B.Sc. and M.Sc. in Electrical Engineering from the Technion-Israel Institute of Technology in 2006, 2006 and 2012 respectively. He has received a Ph.D. in Electrical Engineering and Computer Science from the Massachusetts Institute of Technology in 2018.

He is Currently a Software Engineer with Lyft Inc., working on machine perception systems for Autonomous vehicles. His interests include statistical signal processing, machine learning and efficient sampling strategies.

Stefanie Jegelka received her Diploma in Computer Science (Bioinformatics) with distinction from the University of Tübingen, Germany, in 2007, and her Ph.D. in Computer Science from ETH Zurich, Switzerland, and the Max Planck Institute for Intelligent Systems, Germany, in 2012.

Currently, she is an X-Consortium Career Development Assistant Professor in the Department of EECS at MIT in Cambridge, MA. From 2012 to 2014, she was as a postdoctoral researcher at the University of California, Berkeley. She has also been a visiting researcher at the Simons Institute for the Theory of Computing, Microsoft Research, and Georgetown University. Her current research interests span theory and practice of algorithmic machine learning, in particular reliable learning and optimization with combinatorial structure.

Dr. Jegelka is a member of the ACM and SIAM. She has received a Sloan Research Fellowship in 2018, an NSF CAREER Award in 2016, a DARPA Young Faculty Award in 2017, the German Pattern Recognition Award (Mustererkennungspreis) in 2015 and a Best Paper Award at the International Conference on Machine Learning in 2013.

Gregory W. Wornell (S'83–M'91–SM'00–F'04) received the B.A.Sc. degree from the University of British Columbia, Canada, and the S.M. and Ph.D. degrees from the Massachusetts Institute of Technology, all in electrical engineering and computer science, in 1985, 1987 and 1991, respectively.

Since 1991 he has been on the faculty at MIT, where he is the Sumitomo Professor of Engineering in the Department of Electrical Engineering and Computer Science. At MIT he leads the Signals, Information, and Algorithms Laboratory within the Research Laboratory of Electronics. He is also chair of Graduate Area I (information and system science, electronic and photonic systems, physical science and nanotechnology, and bioelectrical science and engineering) within the EECS department's doctoral program. He has held visiting appointments at the former AT&T Bell Laboratories, Murray Hill, NJ, the University of California, Berkeley, CA, and Hewlett-Packard Laboratories, Palo Alto, CA.

His research interests and publications span the areas of signal processing, information theory, digital communication, statistical inference, and information security, and include architectures for sensing, learning, computing, communication, and storage; systems for computational imaging, vision, and perception; aspects of computational biology and neuroscience; and the design of wireless networks. He has been involved in the Information Theory and Signal Processing societies of the IEEE in a variety of capacities, and maintains a number of close industrial relationships and activities. He has won a number of awards for both his research and teaching, including the 2019 IEEE Leon K. Kirchmayer Graduate Teaching Award.