Optimal Sequential Energy Allocation for Inverse Problems

Raghuram Rangarajan, *Student Member, IEEE,* Raviv Raich, *Member, IEEE,* and Alfred O. Hero III, *Fellow, IEEE*

***Abstract*—This paper investigates the advantages of adaptive waveform amplitude design for estimating parameters of an unknown channel/medium under average energy constraints. We present a statistical framework for sequential design (e.g., design of waveforms in adaptive sensing) of experiments that improves parameter estimation (e.g., unknown channel parameters) performance in terms of reduction in mean-squared error (MSE). We treat an N time step design problem for a linear Gaussian model where the shape of the N input design vectors (one per time step) remains constant and their amplitudes are chosen as a function of past measurements to minimize MSE. For $N = 2$, we derive the optimal energy allocation at the second step as a function of the first measurement. Our adaptive two-step strategy yields an MSE improvement of at least 1.65dB relative to the optimal non-adaptive strategy, but is not implementable since it requires knowledge of the noise amplitude. We then present an implementable design for the two-step strategy which asymptotically achieves optimal performance. Motivated by the optimal two-step strategy, we propose a suboptimal adaptive N-step energy allocation strategy that can achieve an MSE improvement of more than 5dB for $N = 50$. We demonstrate our general approach in the context of MIMO channel estimation and inverse scattering problems.**

***Index Terms*—Parameter estimation, sequential design, energy management, maximum likelihood, MMSE, inverse scattering, channel estimation.**

I. Introduction

Adaptive sensing has been an important topic of research for at least a decade. Many of the classical problems in statistical signal processing such as channel estimation, radar imaging, target tracking, and detection can be presented in the context of adaptive sensing. One of the important components in these adaptive sensing problems is the need for energy management. Most applications are limited by peak power or average power. For example, in sensor network applications, sensors have limited battery life and replacing them is expensive. Safety limits the peak transmit power in medical imaging problems. Energy is also a critical resource in communication systems where reliable communication is necessary at low signal-to-noise ratios. Hence it is important to consider energy limitations in waveform design problems. Most of the effort in previous research has focussed on waveform design under peak power constraints, e.g., sensor management. There has been little effort in developing adaptive waveform design strategies that allocate different amounts of energy to the waveforms over time. Our goal in this paper is to perform waveform amplitude design for adaptive sensing in order to estimate the set of unknown channel parameters or scattering coefficients under an average energy constraint. We formulate this problem as an experimental design problem in the context of sequential parameter estimation. We explain the methodology of experimental design, derive optimal designs, and show performance gains over non-adaptive design techniques. As a final step, we describe in detail how some applications of adaptive sensing such as channel estimation and radar imaging can be cast into this experimental design setting thereby leading to attractive performance gains compared to current literature. Next, we present a review of waveform design and sequential estimation literature to provide a context for our work.

Note: The term 'sequential' is used in different contexts in the literature. In this paper, 'sequential' means that at every time instant, the best signal to transmit is selected from a library that depends on past observations.

A. Related Work - Waveform Design

Early work in waveform design focussed on selecting among a small number of measurement patterns [1]. Radar signal design using a control theoretic approach subject to both average and peak power constraints was addressed in [2] and [3]. The design was non-adaptive and the optimal continuous waveforms were shown to be on-off measurement patterns alternating between zero and peak power levels for a tracking example. In our design, the energy allocation to the waveforms over time are optimally chosen from a continuum of values. Parameterized waveform selection for dynamic state estimation was explored in [4] and [5] where the shape of the waveforms were allowed to vary under constant transmit power. Closed-form solutions to the parameter selection problem were found for a very restrictive set of cases such as one-dimensional target motions. More recently a dynamic waveform selection algorithm for tracking using a class of generalized chirp signals was presented in [6]. In contrast to these efforts, we focus our work in finding optimal waveform amplitudes under an average energy constraint for static parameter estimation. Sensor scheduling can be thought of as an adaptive waveform design problem under a peak power constraint [7] where the goal is to choose the best sensor at each time instant to provide the next measurement. The optimal sensor schedule can be determined a priori and independent of measurements for the case of linear Gaussian systems [8], [9]. The problem of optimal scheduling for the case of hidden Markov model systems was addressed in [10].

This research was partially supported by ARO-DARPA MURI Grant #DAAD19-02-1-0262.

The authors are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109, USA (email: {rangaraj,ravivr,hero }@eecs.umich.edu)

Literature	Type of parameters				Type of design		Type of constraint			Type of control			
	D	R	LSD	NLSD	SQ	NSQ	EN	SN	NONE	EN	WV	SN	NONE
Waveform design [1]			✓	✓	✓				✓		✓		
Sensor scheduling [8]–[10]			✓	✓	✓			✓				✓	
Sequential estimation [11], [12]	✓				✓				✓				✓
Schweppe's design [2], [3]			✓			✓	✓			✓			
RLS [13]	✓				✓				✓				✓
Stein estimator [14], [15]	✓					✓			✓				✓
Kalman filter [16]			✓		✓				✓				✓
Our sequential approach	✓				✓		✓			✓			

TABLE I
KEY TO THE TABLE: D-DETERMINISTIC, R-RANDOM, LSD-LINEAR STATE DYNAMICS, NLSD-NON LINEAR STATE DYNAMICS, SQ-SEQUENTIAL DESIGN, NSQ-NON SEQUENTIAL DESIGN, EN-ENERGY, SN-SENSORS, WV-WAVEFORM PARAMETERS.

In table I, we compare our work with existing literature via different categories.

B. Related Work - Sequential Design for Estimation

The concept of sequential design has been studied by statisticians for many decades [17]–[22] and has found applications in statistics, engineering, biomedicine, and economics. Sequential analysis has been used to solve important problems in statistics such as change-point detection [23], [24], point and interval estimation [25], multi-armed bandit problems [26], quality control [27], sequential testing [28], and stochastic approximation [29]. Robbins pioneered the statistical theory of sequential allocation in his seminal paper [26]. Early research on the application of sequential design to problems of estimation was limited to finding asymptotically risk-efficient point estimates and fixed-width confidence intervals [11], [12], [30], i.e., sequential design was used to solve problems in which a conventional estimate, based on a sample whose size is determined by a suitably chosen stopping rule, achieves certain properties such as bounded risk. For the problem of estimating the mean under unknown variance, it was shown that a sequential two-step method guaranteed specified precision [23], [31], [32], which is not possible using a fixed sample. The statistical sequential design framework assumes a fixed measurement setup while acquiring the data and does not consider energy constraints. In this paper, we adaptively design input parameters to alter the measurement patterns under an average energy constraint to obtain performance gains over non-adaptive strategies.

Another class of problems in sequential estimation is online estimation, where fast updating of parameter estimates are made in real time, called recursive identification in control theory, and adaptive estimation in signal processing. For example, consider the problem of estimating parameter $\boldsymbol{\theta}$ in the following model

$$y_i = \mathbf{x}_i^T \boldsymbol{\theta} + w_i, \quad i = 1, 2, \ldots, n,$$

where $\{\mathbf{x}_i\}$ are the sequence of inputs to the system, $\{w_i\}$ are independent identically distributed (i.i.d) Gaussian random variables with zero mean and $\{y_i\}$ are the set of received signals. The maximum likelihood estimate of $\boldsymbol{\theta}$ is given by the least squares (LS) solution, $\hat{\boldsymbol{\theta}}_{\text{LS}} = \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T\right)^{-1} \left(\sum_{i=1}^n \mathbf{x}_i y_i\right)$. One way of computing the LS estimate is the recursive least squares approach (RLS) [13] which can be written as

$$\begin{aligned} \hat{\boldsymbol{\theta}}_n &= \hat{\boldsymbol{\theta}}_{n-1} + \mathbf{P}_n \mathbf{x}_n (y_n - \mathbf{x}_n^T \hat{\boldsymbol{\theta}}_{n-1}) \\ \mathbf{P}_n &= \mathbf{P}_{n-1} - \frac{\mathbf{P}_{n-1} \mathbf{x}_n \mathbf{x}_n^T \mathbf{P}_{n-1}}{1 + \mathbf{x}_n^T \mathbf{P}_{n-1} \mathbf{x}_n}, \end{aligned}$$

where $\mathbf{P}_n = \left(\sum_{i=1}^n \mathbf{x}_i \mathbf{x}_i^T\right)^{-1}$. The recursive process avoids the computational complexity of inverting the matrix.

In the above formulation it was assumed that the input sequence $\{\mathbf{x}_i\}$ remains fixed. The problem of waveform design is relevant when input $\mathbf{x}_i$ can be adaptively chosen based on the past measurements $y_1, \ldots, y_{i-1}$. Measurement-adaptive estimation has application to a wide variety of areas such as communications and control, medical imaging, radar systems, system identification, and inverse scattering. By measurement-adaptive estimation we mean that one has control over the way measurements are made, e.g., through the selection of waveforms, projections, or transmitted energy. The standard solution for estimating parameters from adaptive measurements is the maximum likelihood (ML) estimator. For the case of classic linear Gaussian model, i.e., a Gaussian observation with unknown mean and known variance, it is well-known [16] that the ML estimator is unbiased and achieves the unbiased Cramér Rao lower bound (CRB). Many researchers have looked at improving the estimation of these parameters by adding a small bias to reduce the MSE. Stein showed that this leads to better estimators that achieve lower MSE than the ML estimator for estimating the mean in a multivariate Gaussian distribution with dimension greater than two [14], [15]. Other alternatives such as the shrinkage estimator [33], Tikhonov regularization [34] and covariance shaping least squares (CSLS) estimator [35] have also been proposed in the literature. While these pioneering efforts present interesting approaches to improve static parameter estimation performance by introducing bias, none of them incorporate the notion of sequential design of input parameters. Our adaptive design of inputs effectively adds bias to achieve reduction in MSE.

In this paper, we formulate a problem of sequentially select-

ing waveform amplitudes for estimating parameters of a linear Gaussian channel model under an average energy constraint over the waveforms and over the number of transmissions. In Section II, the problem of experimental design [36], [37] for sequential parameter estimation is outlined and the analogy between this problem and the waveform design problem is explained. In Section III, closed-form expressions for the optimal design parameters and the corresponding minimum MSE in the single parameter case are derived for a two-step procedure (two time steps). Since the optimal solution requires the knowledge of parameters to be estimated, it is shown in Section IV that the performance of this omniscient solution can be achieved with a parameter independent strategy. In Section IV-A, we describe an N-step sequential energy allocation procedure, which yields more than 5dB gain over non-adaptive methods. These results are extended to the vector parameter case in Section V. Finally in Section VI, we show the applicability of this framework for channel estimation and radar imaging problems.

II. Problem Statement

We begin by introducing nomenclature commonly used throughout the paper. We denote vectors in $\mathbb{C}^M$ by boldface lowercase letters and matrices in $\mathbb{C}^{M\times N}$ by boldface uppercase letters. The identity matrix is denoted by $\mathbf{I}$. We use $(\cdot)^T$ and $(\cdot)^H$ to denote the transpose and conjugate transpose operators, respectively. We denote the l_2-norm of a vector by $\|\cdot\|$, i.e., $\|\mathbf{x}\| = \sqrt{\mathbf{x}^H\mathbf{x}}$. A circularly symmetric complex Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C}$ is denoted as $\mathcal{CN}(\boldsymbol{\mu},\mathbf{C})$. $\mathrm{E}[\cdot]$ and $\mathrm{tr}(\cdot)$ denote the statistical expectation and trace operators, respectively. The terms MSE and SNR are abbreviations to mean-squared error and signal-to-noise ratio, respectively.

Let $\boldsymbol{\theta} = [\theta_1,\ldots,\theta_M]^T$ be the M-element vector of unknown parameters. The problem of estimating $\boldsymbol{\theta}$ in noise can then be written as

$$\mathbf{y}_i = \mathbf{f}(\mathbf{x}_i,\boldsymbol{\theta}) + \mathbf{n}_i, \quad i = 1,2,\ldots,N, \tag{1}$$

where $\{\mathbf{n}_i\}$ is an i.i.d. random process corrupting the function of the parameters of interest $\mathbf{f}(\mathbf{x}_i,\boldsymbol{\theta})$ and i denotes the time index. The T-element design parameter vectors, $\{\mathbf{x}_i\}_{i=1}^N$ can depend on the past measurements: $\mathbf{x}_i = \mathbf{x}_i(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1})$, where $\mathbf{y}_i$ is the i^{th} K-element observation vector. In the context of adaptive sensing, $\mathbf{f}(\mathbf{x}_i,\boldsymbol{\theta})$ represents the response of the medium, T and K denote the number of transmit and receive antennas respectively, $\{\mathbf{x}_i\}_{i=1}^N$ are the set of waveforms to be designed, $\boldsymbol{\theta}$ are the set of channel parameters or scattering coefficients to be estimated using the set of received signals $\{\mathbf{y}_i\}_{i=1}^N$. For the classic estimation problem in a linear Gaussian model, we have $\mathbf{f}(\mathbf{x}_i,\boldsymbol{\theta}) = \mathbf{H}(\mathbf{x}_i)\boldsymbol{\theta}$, $\mathbf{H}(\mathbf{x}_i) = [\mathbf{h}_1(\mathbf{x}_i),\mathbf{h}_2(\mathbf{x}_i),\ldots,\mathbf{h}_M(\mathbf{x}_i)]$ is a known $K\times M$ matrix and linear in $\mathbf{x}_i$ and $\mathbf{n}_i$ is a $\mathcal{CN}(\mathbf{0},\sigma^2\mathbf{I})$ random vector. When $\mathbf{H}(\mathbf{x})$ is linear in $\mathbf{x}$, we can write $\mathbf{h}_l(\mathbf{x}) = \boldsymbol{H}_l\mathbf{x}, l = 1,2,\ldots,M$. In this case $\mathbf{H}(\cdot)$ is uniquely determined by the matrices $\{\boldsymbol{H}_1,\boldsymbol{H}_2,\ldots,\boldsymbol{H}_M\}$. The linear Gaussian model has been widely adopted in many studies [38], [39] including channel estimation [40] and radar imaging [41] problems. The set of observations for the case of a scalar parameter θ_1 are

$$\mathbf{y}_i = \mathbf{h}_1(\mathbf{x}_i)\theta_1 + \mathbf{n}_i, \quad i = 1,2,\ldots,N. \tag{2}$$

An N-step design procedure specifies a sequence of functions $\{\mathbf{x}_i(\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_{i-1})\}_{i=1}^N$ corresponding to the N transmitted signal waveforms after receiving the previous measurements. An optimal N-step procedure selects the design vectors so that the MSE of the maximum likelihood (ML) estimator, $\hat{\theta}_1^{(N)}(\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_N)$ is minimized subject to the average energy constraint, $\mathrm{E}\left[\sum_{i=1}^N \|\mathbf{x}_i\|^2\right] \le E_0$, where E_0 is the total available energy. The ML estimator of θ_1 for the N-step procedure is given by

$$\hat{\theta}_1^{(N)} = \frac{\sum_{i=1}^N \mathbf{h}_1(\mathbf{x}_i)^H\mathbf{y}_i}{\sum_{i=1}^N \|\mathbf{h}_1(\mathbf{x}_i)\|^2} \tag{3}$$

and the corresponding $\mathrm{MSE}\left(\{\mathbf{x}_i\}_{i=1}^N\right) \triangleq \mathrm{E}\left[\left|\hat{\theta}_1^{(N)} - \theta_1\right|^2\right]$ is

$$\mathrm{MSE}^{(N)}\left(\{\mathbf{x}_i\}_{i=1}^N\right) = \mathrm{E}\left[\left|\frac{\sum_{i=1}^N \mathbf{h}_1(\mathbf{x}_i)^H\mathbf{n}_i}{\sum_{i=1}^N \|\mathbf{h}_1(\mathbf{x}_i)\|^2}\right|^2\right]. \tag{4}$$

Denote $E_i(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1}) = \|\mathbf{x}_i(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1})\|^2$, where $E_i(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1})$ represents the energy allocated to each time step i. Define $\mathcal{E}\left[\{\mathbf{x}_i(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1})\}_{i=1}^N\right]$ as the average energy in the design parameters for the N-step procedure. The average energy constraint can be written as

$$\mathcal{E}\left[\{\mathbf{x}_i(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1})\}_{i=1}^N\right] = \mathrm{E}\left[\sum_{i=1}^N \|\mathbf{x}_i\|^2\right] \le E_0. \tag{5}$$

Our goal is to find the best sequence of the design vectors $\{\mathbf{x}_i\}_{i=1}^N$ to minimize the $\mathrm{MSE}^{(N)}\left(\{\mathbf{x}_i\}_{i=1}^N\right)$ in (4) under the average energy constraint in (5).

A. Non-adaptive strategy

As a benchmark for comparison, we consider the non-adaptive case where $\mathbf{x}_i(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1})$ is deterministic, independent of $\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_{i-1}$, $\|\mathbf{x}_i\|^2 = E_i$, and $\sum_{i=1}^N E_i \le E_0$. Simplifying the expression for MSE in (4), we have

$$\mathrm{MSE}^{(N)} = \frac{\sigma^2}{\sum_{i=1}^N \|\mathbf{h}_1(\mathbf{x}_i)\|^2} \ge \frac{\sigma^2}{E_0\lambda_{\mathrm{m}}(\boldsymbol{H}_1)}, \tag{6}$$

where equality is achieved iff $\forall i$ $\mathbf{x}_i \propto \mathbf{v}_{\mathrm{m}}$, the normalized eigenvector corresponding to $\lambda_{\mathrm{m}}(\boldsymbol{H}_1)$, the maximum eigenvalue of the matrix $\boldsymbol{H}_1^H\boldsymbol{H}_1$. Note $\lambda_{\mathrm{m}}(\boldsymbol{H}_1) = \max_{\mathbf{x}}(\mathbf{x}^H\boldsymbol{H}_1^H\boldsymbol{H}_1\mathbf{x})/(\mathbf{x}^H\mathbf{x}) = \max_{\mathbf{x}}\|\mathbf{h}_1(\mathbf{x})\|^2/\|\mathbf{x}\|^2$. Furthermore, the performance of the ML estimator does not depend on the energy allocation. Hence, without loss of generality we can assume all energy is allocated to the first transmission which implies that any N-step non-adaptive strategy is no better than the optimal one-step strategy. We define $\mathrm{SNR}\left(\{\mathbf{x}_i\}_{i=1}^N\right)$ as

$$\mathrm{SNR}^{(N)} = \frac{\lambda_{\mathrm{m}}(\boldsymbol{H}_1)\mathcal{E}\left[\{\mathbf{x}_i(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1})\}_{i=1}^N\right]}{\sigma^2}. \tag{7}$$

Then the average energy constraint in (5) is equivalent to $\mathrm{SNR}^{(N)} \le \mathrm{SNR}_0$, where $\mathrm{SNR}_0 = \lambda_{\mathrm{m}}(\boldsymbol{H}_1)E_0/\sigma^2$. We show

in [42] that the problem of minimizing $\text{MSE}^{(N)}$ subject to $\text{SNR}^{(N)} \leq \text{SNR}_0$ is equivalent to minimizing $\text{MSE}^{(N)} \times \text{SNR}^{(N)}$. Thus we use the two minimization criteria interchangeably in the remainder of this paper. The product of MSE and SNR is

$$\text{MSE}^{(N)} \times \text{SNR}^{(N)} = \text{E}\left[\left|\frac{\sum_{i=1}^{N} \mathbf{h}_1(\mathbf{x}_i)^H \mathbf{n}_i}{\sum_{i=1}^{N} \|\mathbf{h}_1(\mathbf{x}_i)\|^2}\right|^2\right] \times \left(\frac{\lambda_{\text{m}}(\boldsymbol{H}_1)\text{E}\left[\sum_{i=1}^{N}\|\mathbf{x}_i\|^2\right]}{\sigma^2}\right) \quad (8)$$

and the minimum MSE for the one-step (or non-adaptive N-step) strategy satisfies

$$\text{MSE}_{\min}^{(1)} \times \text{SNR}_0 = 1. \quad (9)$$

While our goal is to find optimal input design parameters, $\{\mathbf{x}_j(\mathbf{y}_1, \ldots, \mathbf{y}_{j-1})\}_{j=1}^{N}$ which achieve minimum MSE, any suboptimal design that guarantees $\text{MSE}^{(N)} \times \text{SNR}_0 < 1$ is also of interest. We first look at a two-step sequential design procedure. A word of caution: in Sections III and III-A we develop optimal and suboptimal strategies where the solutions require the knowledge of the unknown parameter θ_1. However, in Section IV we present a θ_1-independent design which asymptotically achieves the performance of the 'omniscient' strategies.

III. Omniscient Optimal Two-step Sequential Strategy

In the two-step sequential procedure, we have $N = 2$ time steps where in each time step $i = 1, 2$, we can control input design parameter $\mathbf{x}_i$ to obtain observation $\mathbf{y}_i$. For a two-step process, we have

$$\mathbf{y}_1 = \mathbf{h}_1(\mathbf{x}_1)\theta_1 + \mathbf{n}_1 \quad (10)$$

$$\mathbf{y}_2 = \mathbf{h}_1(\mathbf{x}_2(\mathbf{y}_1))\theta_1 + \mathbf{n}_2. \quad (11)$$

The ML estimator of θ_1 for a two-step procedure from (3) is

$$\hat{\theta}_1^{(2)} = \frac{\mathbf{h}_1(\mathbf{x}_1)^H\mathbf{y}_1 + \mathbf{h}_1(\mathbf{x}_2)^H\mathbf{y}_2}{\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2} \quad (12)$$

and its MSE from (4) is given by

$$\text{MSE}^{(2)}(\mathbf{x}_1, \mathbf{x}_2) = \text{E}\left[\frac{|\mathbf{h}_1(\mathbf{x}_1)^H\mathbf{n}_1 + \mathbf{h}_1(\mathbf{x}_2)^H\mathbf{n}_2|^2}{(\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2)^2}\right]. \quad (13)$$

We assume that the shape of the optimal designs, i.e., $\{\mathbf{x}_i/\|\mathbf{x}_i\|\}$ is the one-step optimum given by $\mathbf{v}_{\text{m}}$ defined below (6) and minimize the MSE over the energy of the design parameters. Denote $\|\mathbf{x}_1\| = \sqrt{E_0}\alpha_1$ and $\|\mathbf{x}_2(\mathbf{y}_1)\| = \sqrt{E_0}\alpha_2(\mathbf{y}_1)$. Under the sequential design framework, we select

$$\mathbf{x}_1 = \sqrt{E_0}\ \alpha_1\mathbf{v}_{\text{m}} \quad (14)$$

$$\mathbf{x}_2(\mathbf{y}_1) = \sqrt{E_0}\ \alpha_2(\mathbf{y}_1)\mathbf{v}_{\text{m}}, \quad (15)$$

where α_1 and $\alpha_2(\cdot)$ are real-valued scalars. The average energy constraint from (5) can then be written as

$$\text{E}\left[\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)\right] \leq 1. \quad (16)$$

We use Lagrangian multipliers to minimize the MSE in (13) with respect to α_1 and $\alpha_2(\cdot)$ under the energy constraint in (16). The objective function to be minimized can then be written as

$$\text{E}\left[\frac{|\mathbf{h}_1(\mathbf{x}_1)^H\mathbf{n}_1 + \mathbf{h}_1(\mathbf{x}_2)^H\mathbf{n}_2|^2}{(\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2)^2}\right] + \gamma\left(\alpha_1^2 + \text{E}\left[\alpha_2^2(\mathbf{y}_1)\right]\right)$$

In [42], we show that the optimal solution to $\alpha_2(\mathbf{y}_1)$ depends on $\mathbf{y}_1$ only through the function $\tilde{n}_1(\mathbf{y}_1; \theta_1)$, where

$$\tilde{n}_1(\mathbf{y}_1; \theta_1) = \frac{\mathbf{h}_1(\mathbf{v}_{\text{m}})^H(\mathbf{y}_1 - \mathbf{h}_1(\mathbf{x}_1)\theta_1)}{\|\mathbf{h}_1(\mathbf{v}_{\text{m}})\|\sigma} = \frac{\mathbf{h}_1(\mathbf{v}_{\text{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\text{m}})\|}\frac{\mathbf{n}_1}{\sigma} \quad (17)$$

$\sim \mathcal{CN}(0, 1)$. Hence we denote the solution as $\alpha_2\left(\tilde{n}_1(\mathbf{y}_1; \theta_1)\right)$. Let $g\left(\tilde{n}_1(\mathbf{y}_1; \theta_1)\right) = \left(1 + \frac{\alpha_2^2(\tilde{n}_1(\mathbf{y}_1;\theta_1))}{\alpha_1^2}\right)$. Setting the derivative of the objective function with respect to g to zero yields

$$g^3 - \frac{1}{\gamma'}g + 2\frac{1 - |\tilde{n}_1|^2}{\gamma'} = 0, \quad (18)$$

where $\gamma' = \gamma\alpha_1^4\text{SNR}_0$. The function g that minimizes MSE is the root of the third-order polynomial in (18), real-valued and greater than or equal to 1. If more than one real-valued solution greater than 1 to the cubic equation exists, the optimal solution to g will be the root that achieves minimum MSE. The optimal g for every $\tilde{n}_1$ and γ' is denoted by $g_{\gamma'}(\tilde{n}_1)$. Also $\text{E}\left[g_{\gamma'}(\tilde{n}_1)\right] = \frac{1}{\alpha_1^2}$. Therefore, finding α_1 that minimizes MSE is equivalent to finding γ' that minimizes MSE. We obtain $g_{\gamma'}(\tilde{n}_1)$ for every γ' and use a brute force grid search to find the optimal γ' that minimizes the objective function. The MSE is minimized at $\gamma'^* \approx 0.22$, or $\alpha_1^* \approx 0.7421$. The optimal α_2 is given by the relation $\alpha_2^*\left(\tilde{n}_1(\mathbf{y}_1; \theta_1)\right) = \alpha_1^*\sqrt{\left(g_{\gamma'^*}\left(\tilde{n}_1(\mathbf{y}_1; \theta_1)\right) - 1\right)}$. Since this solution depends on the unknown parameter θ_1, we call this minimizer an "omniscient" energy allocation strategy. For the optimal solution, the product of MSE $\times$ SNR is

$$\text{MSE}_{\min}^{(2)} \times \text{SNR}_0 \approx 0.68. \quad (19)$$

This corresponds to a 32% improvement in performance or a 1.67dB gain in terms of SNR for the two-step design when compared to the one-step procedure for which $\text{MSE}_{\min}^{(1)} \times \text{SNR}_0 = 1$.

The optimal energy allocation at the second step, $\alpha_2^{*2}\left(\tilde{n}_1(\mathbf{y}_1; \theta_1)\right)$ as shown in Fig. 1 (solid) is a thresholding function, i.e., α_2^* is zero for $|\tilde{n}_1|^2 \leq 0.59$. This solution implies that when the actual realization of the normalized noise along $\mathbf{h}_1(\mathbf{v}_{\text{m}})$ in the first step is small enough, then the second measurement becomes unnecessary. On the other hand, when the normalized noise along $\mathbf{h}_1(\mathbf{v}_{\text{m}})$ exceeds a threshold, then there is some merit in incorporating the information from the second measurement. The solution also suggests that the higher the noise magnitude at the first step, the more the energy that needs to be used. However, the probability of allocating energy greater than a particular value decreases exponentially with that energy value. Nevertheless in applications with a peak energy constraint, the transmission of the optimal energy at the second stage may not always

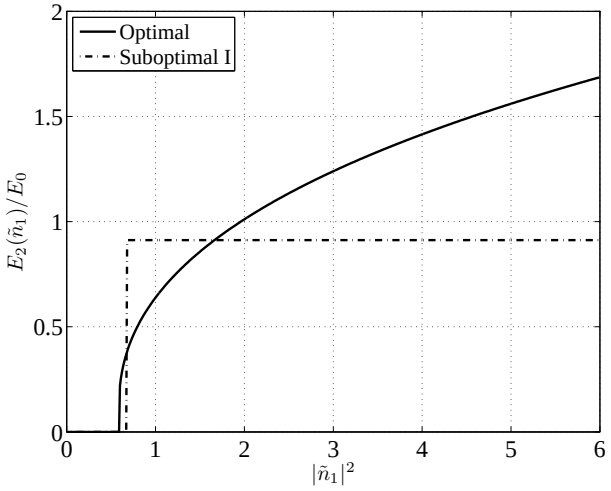


Fig. 1. Plot of the optimal and suboptimal solution to the normalized energy transmitted at the second stage as functions of received signal at first stage.

be possible. Hence, in the following subsection we look at a suboptimal solution which takes into account this constraint and still achieves near optimal performance.

A. Omniscient Suboptimal Two-step Strategy

The optimal solution in Section III is a thresholding function, where energy allocated to the second stage is zero if the noise magnitude at the first step is less than a threshold and increases with increasing noise magnitudes otherwise. For the suboptimal solution, we use a binary energy allocation strategy at the second stage based on the noise magnitude at the first step, i.e., we allocate a fixed nonzero energy if the noise magnitude is greater than a threshold else we allocate zero energy. The suboptimal solution to the design vectors $\mathbf{x}_1$ and $\mathbf{x}_2$ is then of the form

$$\mathbf{x}_1 = \mathbf{v}_{\mathrm{m}} \sqrt{E_0}\, \alpha_1 \quad (20)$$

$$\mathbf{x}_2 = \mathbf{v}_{\mathrm{m}} \sqrt{E_0}\, \alpha_2\, I\left(|\tilde{n}_1|^2 > \rho\right), \quad (21)$$

where $\tilde{n}_1$ is defined in (17), α_1, α_2 are design parameters independent of $\mathbf{y}_1$ and $I(\cdot)$ is the indicator function, i.e.,

$$I(A) = \begin{cases} 1, & A \text{ is true} \\ 0, & A \text{ is false.} \end{cases}$$

The SNR of the suboptimal two-step procedure is

$$\mathrm{SNR}^{(2)} = \mathrm{SNR}_0 \left(\alpha_1^2 + \alpha_2^2 P\left(|\tilde{n}_1|^2 > \rho\right)\right). \quad (22)$$

The MSE of the ML estimator under this suboptimal solution using (13) is

$$\mathrm{MSE}^{(2)} = \frac{1}{\mathrm{SNR}_0}\mathrm{E}\left[\frac{\alpha_1^2|\tilde{n}_1|^2 + \alpha_2^2}{\left(\alpha_1^2+\alpha_2^2\right)^2} I\left(|\tilde{n}_1|^2 \geq \rho\right)\right] + \frac{1}{\mathrm{SNR}_0}\mathrm{E}\left[\frac{|\tilde{n}_1|^2}{\alpha_1^2} I\left(|\tilde{n}_1|^2 < \rho\right)\right]. \quad (23)$$

Denote $\beta = \frac{\alpha_1^2}{\alpha_1^2+\alpha_2^2}$, $0 \leq \beta \leq 1$. Substituting for β in the expressions for $\mathrm{MSE}^{(2)}$ and $\mathrm{SNR}^{(2)}$ in (23) and (22), and using the fact that $\mathrm{E}\left[I(|x|^2 \geq \rho)\right] = e^{-\rho}$ and $\mathrm{E}\left[|x|^2 I(|x|^2 \geq \rho)\right] = \rho e^{-\rho}$ when $x \sim \mathcal{CN}(0,1)$, the expression $\mathrm{MSE}^{(2)} \times \mathrm{SNR}^{(2)}$ simplifies to

$$\mathrm{MSE}^{(2)} \times \mathrm{SNR}^{(2)} = ((\beta\rho+1)e^{-\rho} + \frac{1}{\beta}(1-(1+\rho)e^{-\rho})) \times \left(\beta + (1-\beta)e^{-\rho}\right). \quad (24)$$

Minimizing $\mathrm{MSE}^{(2)} \times \mathrm{SNR}^{(2)}$ with respect to β and ρ through a grid search for $\beta \in [0,1]$ and $\rho \in [0,\infty)$ yields $\beta^* \approx 0.37$, $\rho^* \approx 0.675$. It follows that $\alpha_1^* \approx 0.7319$ and $\alpha_2^* \approx 0.9550$. Substituting for the optimal values of $\alpha_1^*, \alpha_2^*, \beta^*, \rho^*$ in (24) and (22), and simplifying yields

$$\mathrm{MSE}^{(2)}_{\min} \times \mathrm{SNR}_0 \approx 0.7143. \quad (25)$$

This translates to a 28.47% improvement in MSE performance or a 1.5dB savings in terms of SNR. The suboptimal solution to the energy design is shown in Fig. 1 by a dashed dotted line indicated as Suboptimal-I. Thus, while the suboptimal strategy limits the peak transmit power to $\max\left(\alpha_1^{*2}, \alpha_2^{*2}\right) E_0$, it is able to achieve near optimal performance.

In the previous section, we addressed the problem of minimizing MSE subject to an average energy constraint, $\mathrm{E}\left[\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2\right] \leq E_0$. An average energy constraint implies that the total allocated energy averaged over repeated trials of the two-step experiment is constrained to be less than or equal to E_0. This is less restrictive than the strict energy constraint $\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 \leq E_0$, as any solution satisfying this constraint satisfies the average energy constraint but not vice versa. The problem of minimizing the MSE in (13) under this strict energy constraint was addressed in the context of radar imaging in [43]. We show in [42] that the optimal two-step design under the strict constraint is given by

$$\mathbf{x}_1 = \alpha_1^* \sqrt{E_0}\mathbf{v}_{\mathrm{m}}$$

$$\mathbf{x}_2 = \alpha_2^* \sqrt{E_0}\, \mathbf{v}_{\mathrm{m}}\, I\left(|\tilde{n}_1|^2 > \rho^*\right),$$

where $\alpha_1^* \approx 0.628$, $\alpha_2^* \approx 0.7782$, and $\rho^* \approx 0.2831$. The minimum MSE is then given by $\mathrm{MSE}^{(2)}_{\min} \times \mathrm{SNR}_0 \approx 0.9219$. The optimal solution satisfies the strict energy constraint with equality but the average energy used is $E_0(\alpha_1^{*2} + \alpha_2^{*2} e^{-\rho^*}) \approx 0.8550 E_0$. The solution to the two-step strategy under this strict energy constraint can also be derived by imposing an additional constraint, $\alpha_1^2 + \alpha_2^2 \leq 1$ to the suboptimal design problem described earlier in this section. In the following section, we design a θ_1-independent design strategy that achieves the optimal performance asymptotically and allows for any peak power constraint in the design.

IV. Parameter Independent Two-step Design Strategy

Consider the optimal design for the two-step procedure

$$\mathbf{x}_1 = \sqrt{E_0}\, \alpha_1^* \mathbf{v}_{\mathrm{m}}$$

$$\mathbf{x}_2 = \sqrt{E_0}\, \alpha_2^*(\tilde{n}_1(\mathbf{y}_1;\theta_1))\mathbf{v}_{\mathrm{m}}$$

$$= \sqrt{E_0}\, \alpha_2^* \left(\left|\frac{\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\|} \frac{(\mathbf{y}_1 - \sqrt{E_0}\alpha_1^* \mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\theta_1)}{\sigma}\right|\right) \mathbf{v}_{\mathrm{m}}. \quad (26)$$

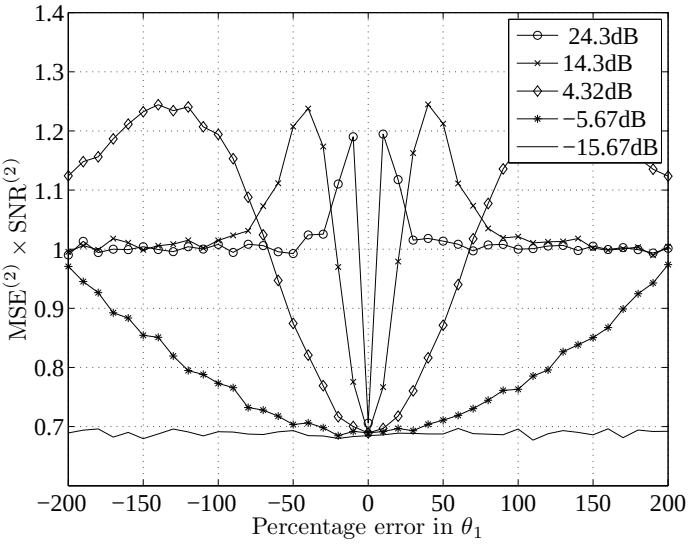


Fig. 2. Plot of reduction in MSE versus percentage error in the guess of parameter θ_1 for various SNR.

We showed that by designing α_1 and α_2 optimally we can gain up to 32% improvement in estimator performance. But the "omniscient" solution (26) depends on the parameter to be estimated. Here, we prove that we can approach the optimal two-step gain by implementing a θ_1-independent energy allocation strategy when θ_1 is bounded, i.e., $\theta_1 \in [\theta_{\min}, \theta_{\max}]$, $\theta_{\min}, \theta_{\max} \in \mathbb{R}$. We describe the intuition behind the proposed solution in this section. The details of the proof can be found in [42]. Since we do not know the value of the actual parameter, we replace θ_1 by a 'guess' of θ_1, say θ_g, in the optimal solution to the design at the second step given in (26). The resulting suboptimal design is

$$\mathbf{x}_1 = \sqrt{E_0}\alpha_1^*\mathbf{v}_{\mathrm{m}} \quad (27)$$

$$\mathbf{x}_2 = \sqrt{E_0}\alpha_2^*\left(\left|\frac{\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\|}\frac{(\mathbf{y}_1 - \sqrt{E_0}\alpha_1^*\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\theta_g)}{\sigma}\right|\right)\mathbf{v}_{\mathrm{m}}$$
$$= \sqrt{E_0}\alpha_2^*\left(|\tilde{n}_1 + z|\right)\ \mathbf{v}_{\mathrm{m}}, \quad (28)$$

where

$$z = \frac{\alpha_1^*\sqrt{E_0}\|\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\|}{\sigma}(\theta_1 - \theta_g) = \alpha_1^*\sqrt{\mathrm{SNR}_0}(\theta_1 - \theta_g) \quad (29)$$

and $\tilde{n}_1$, which is defined in (17) is $\mathcal{CN}(0,1)$. Substituting the above suboptimal solution in the expression for $\mathrm{MSE}^{(N)} \times \mathrm{SNR}^{(N)}$ in (8) and simplifying, we obtain

$$\eta(z) = \mathrm{MSE}^{(2)}(z) \times \mathrm{SNR}^{(2)}(z)$$
$$= \mathrm{E}\left[\frac{\alpha_1^{*2}|\tilde{n}_1|^2 + \alpha_2^{*2}(|\tilde{n}_1 + z|)}{(\alpha_1^{*2} + \alpha_2^{*2}(|\tilde{n}_1 + z|))^2}\right]$$
$$\times \mathrm{E}\left[\alpha_1^{*2} + \alpha_2^{*2}(|\tilde{n}_1 + z|)\right]. \quad (30)$$

Figure 2 shows $\eta(z)$ in (30) as a function of the percentage error in the guess of θ_1, $100\left(\frac{\theta_1 - \theta_g}{\theta_1}\right)$ for varying SNR_0. The plot indicates that when $\theta_g = \theta_1$, the optimal performance of the adaptive two-step strategy is achieved for all SNR. At high SNR, for certain values of $|\theta_1 - \theta_g|$, the two-step strategy defined by equations (27) and (28) performs worse than a single step strategy with signal-to-noise ratio SNR_0. This is because the solution presented in (27) and (28) in terms of scalar α_1^* and thresholding function $\alpha_2^*(\cdot)$ were optimized for $\tilde{n}_1 + z \sim \mathcal{CN}(0,1)$, i.e., when $z = 0$. When $\theta_g \neq \theta_1$, the following happens: $z \neq 0$, $\tilde{n}_1 + z \sim \mathcal{CN}(z,1)$, and the design parameters α_1^* and $\alpha_2^*(\cdot)$, which were found optimally for $\tilde{n}_1 + z \sim \mathcal{CN}(0,1)$ ($z = 0$) are no longer optimal. When $|\theta_1 - \theta_g|$ is large, z in (29) is a large constant and hence $\tilde{n}_1$ is a negligible term compared to z with high probability. In other words, $\alpha_2^*(\tilde{n}_1 + z)$ can be made arbitrarily close to $\alpha_2^*(z)$ with high probability as z tends to infinity. This implies that the strategy becomes equivalent to a two-step non-adaptive strategy with a specific non-adaptive energy distribution between the two steps whose performance is given by $\mathrm{MSE}^{(2)} \times \mathrm{SNR}_0 = 1$ from Section II-A. Thus we observe that the performance of the two-step strategy tends to 1 for large $|\theta_1 - \theta_g|$.

The optimal solution to $\mathrm{MSE}^{(2)}(z) \times \mathrm{SNR}^{(2)}(z)$ is achieved when $z = 0$. There are two ways that drive $z \to 0$. If $\theta_1 = \theta_g$, then $z = 0$ and we have $\eta(0) = \eta^* = \mathrm{MSE}^{(2)}_{\min} \times \mathrm{SNR}_0 \approx 0.68$, the optimal two-step performance. Since θ_g is arbitrary, $|\theta_1 - \theta_g| > 0$; the two-step design is not optimal and therefore $\mathrm{MSE}^{(2)} \times \mathrm{SNR}_0 = \eta(z) > \eta^*$. The other way to achieve the optimal solution is to make SNR_0 as small as possible. Note that if SNR_0 is sufficiently small $\mathrm{MSE}^{(2)} \times \mathrm{SNR}^{(2)}$ approaches its minimal value. Since $\mathrm{SNR}^{(2)} \leq \mathrm{SNR}_0$, driving the SNR_0 to zero, drives the $\mathrm{MSE}^{(2)}$ to infinity. To overcome this problem, we propose an $N \times 2$-*step procedure* to allow the SNR_0 to be fixed while driving $z \to 0$. The $N \times 2$-step algorithm is outlined in Fig. 3. Any peak power constraint can also be satisfied using the $N \times 2$-step strategy by choosing a sufficiently large N. The most important information in Fig. 2 is the performance of the two-step strategy under the low SNR regime since each 2-step procedure in the $N \times 2$-step strategy works at $(1/N)^{\text{th}}$ of the total SNR. Hence as N becomes large, SNR in each experiment is very small and the lack of knowledge of θ_1 plays a negligible effect on the performance as z is made close to zero through the SNR factor.

A. Design of N-step procedure

In Sections III and IV, we derived the omniscient optimal two-step design to minimize the MSE and proved that the optimal performance can be achieved asymptotically using an $N \times 2$-step strategy. But the $N \times 2$-step strategy is a specific case of a $2N$-step design. In this subsection, we generalize the suboptimal solution from the 2-step case to the N-step case as follows: we assume that the shape of the design vector is fixed and look at the energy allocation among the various steps. The set of observations are as defined in (2). Let the shape of the design vector $\mathbf{x}_i$ be $\mathbf{v}_{\mathrm{m}}$ and the energy at step i, $E_i = \alpha_i^2(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})$, i.e., $\mathbf{x}_i = \mathbf{v}_{\mathrm{m}}\alpha_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})$, $1 \leq i \leq N$. Then

$$\alpha_1 = A_1,$$
$$\alpha_i = A_i I\left(\frac{|\sum_{j=1}^{i-1}\mathbf{h}_1(\mathbf{x}_j)^H\mathbf{n}_j|^2}{\sum_{j=1}^{i-1}\|\mathbf{h}_1(\mathbf{x}_j)\|^2\sigma^2} \geq \rho_i\right),\ i \geq 2, \quad (31)$$

where $\{A_i, \rho_i\}$ are design parameters. This approximate solution is motivated from the suboptimal thresholding solution

- *Step 1: Perform N independent two-step suboptimal experiments with inputs $\frac{1}{\sqrt{N}}\mathbf{x}_1$ and $\frac{1}{\sqrt{N}}\mathbf{x}_2$ where $\mathbf{x}_1$ and $\mathbf{x}_2$ are given in (27) and (28) respectively, i.e., use energy E_0/N in each of the N experiments.*
  - The SNR of the $2N$-step procedure is $\text{SNR}^{(2N)}(z) = N\text{SNR}^{(2),1}(z) = \text{SNR}^{(2)}(z/\sqrt{N})$ where $\text{SNR}^{(2),k}$ is the SNR of the k^{th} two-step experiment. The first equality follows from the fact that $\{\text{SNR}^{(2),k}\}_{k=1}^{N}$ are identical as the N experiments are independent while the second equality follows from the fact that each two-step experiment uses only $(1/N)^{\text{th}}$ of the total energy.
- *Step 2: Obtain ML estimate from each step as $\hat{\theta}_1^{(2),k}$ and average the N estimates to obtain the ML estimator of the $N \times 2$-step strategy as $\hat{\theta}_1^{(2N)} = \frac{1}{N}\sum_{k=1}^{N}\hat{\theta}_1^{(2),k}$.*
  - The MSE of $\hat{\theta}_1^{(2N)}$ is given by $\text{MSE}^{(2N)}(z) = \frac{1}{N}\text{MSE}^{(2),1}(z) = \text{MSE}^{(2)}(z/\sqrt{N})$, where $\text{MSE}^{(2),k}$ is the MSE of each two-step estimator $\hat{\theta}_1^{(2),k}$. The first equality follows from the fact that $\{\text{MSE}^{(2),k}\}_{k=1}^{N}$ are identical as the N experiments are independent while the second equality follows from the fact that each two-step experiment uses only $(1/N)^{\text{th}}$ of the total energy.
- From Steps 1 and 2, we have $\text{MSE}^{(2N)}(z) \times \text{SNR}^{(2N)}(z) = \text{MSE}^{(2)}(z/\sqrt{N})\text{SNR}^{(2)}(z/\sqrt{N})$. As $N \to \infty$, $z/\sqrt{N} \to 0$ and $\text{MSE}^{(2N)}(z) \times \text{SNR}^{(2N)}(z) \to \eta^*$, i.e., minimal MSE is achieved. The details of the proof can be found in [42].

Fig. 3. Description of the $N\times$ two-step procedure.

to the two-step case derived in Section III-A. Note that the definition of the amplitudes at each stage is recursive, i.e., the amplitude design α_i depends on past inputs $\mathbf{x}_1, \ldots, \mathbf{x}_{i-1}$ which in turn depends on $\alpha_1, \ldots, \alpha_{i-1}$. To simplify our analysis, we make the assumption $\rho_1 \leq \rho_2 \leq \ldots \leq \rho_N$. Then,

$$\alpha_2 = A_2 I\left(\left|\frac{\mathbf{h}_1(\mathbf{x}_1)^H}{\|\mathbf{h}_1(\mathbf{x}_1)\|}\frac{\mathbf{n}_1}{\sigma}\right|^2 \geq \rho_2\right) = A_2 I\left(|\tilde{n}_1|^2 \geq \rho_2\right), \tag{32}$$

where $\tilde{n}_i = \frac{\mathbf{h}_1(\mathbf{v}_\text{m})^H}{\|\mathbf{h}_1(\mathbf{v}_\text{m})\|}\frac{\mathbf{n}_i}{\sigma}$ are i.i.d $\mathcal{CN}(0,1)$ random variables. Following the same procedure, a general expression for α_i can be written as

$$\alpha_i = A_i \prod_{s=1}^{i-1} I\left(|w_s|^2 \geq \rho_{s+1}\right), \tag{33}$$

where $w_s = \frac{\sum_{j=1}^{s} A_j \tilde{n}_j}{\sqrt{\sum_{j=1}^{s}|A_j|^2}}$. This form states that the stopping criteria at time step s is when the magnitude of the average noise, w_s drops below the threshold ρ_{s+1}. The goal is to minimize $\mathcal{G}_N = \text{MSE}^{(N)} \times \text{SNR}^{(N)}$ of this N-step procedure with respect to $\mathbf{A} = [A_1, \ldots, A_N]$ and $\boldsymbol{\rho} = [\rho_1, \ldots, \rho_N]$.

Since there is no closed-form solution to this $2N$ dimensional optimization [42], we evaluate the performance of suboptimal solutions to the design vectors $\mathbf{A}$ and $\boldsymbol{\rho}$. For our simulations, we choose $\rho_i = (i-1)/(N-1)\,\rho_{\max}$, $1 \leq i \leq N$. Furthermore, we choose $\mathbf{A}$ as $\{A_i = d\ \alpha_1^*,\ \text{odd } i;\ A_i = d\ \alpha_2^*,\ \text{even } i\}$, where α_1^*, α_2^* are optimal values from the suboptimal solution presented in Section III-A and d is chosen to satisfy the average energy constraint. The intuition behind this choice of $\mathbf{A}$ and $\boldsymbol{\rho}$ is motivated through an asymptotic result derived in [42]. We evaluate the performance of the N-step procedure with these parameters through theory and verify the theory using simulations. Performance gains, $\mathcal{G}_N$ (in dB) are presented in Fig. 4. By designing this N-step procedure, we are essentially altering the Gaussian statistics of the measurement noise to obtain improvements in performance. In Fig. 5, we illustrate how the distribution of the estimation residuals changes with the number of the steps. We see that

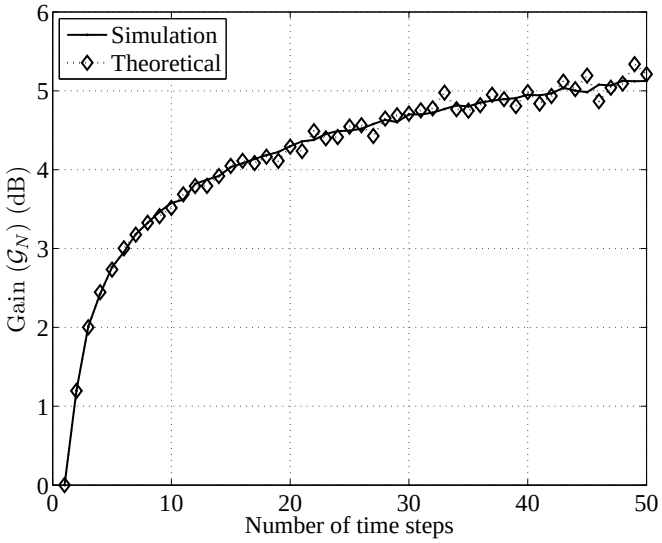


Fig. 4. Plot of gains obtained through suboptimal N-step procedure as a function of N through theory and simulations.

in 50 steps, we are able to achieve gains of more than 5dB. In Section IV, we showed that the two-step gain can be achieved using an $N \times 2$-step strategy, i.e., in $2N$ steps. The basic motivating factor was to reduce the SNR in each experiment and achieve the diversity gain by increasing the number of steps. For the general N-step strategy, progressive reduction in SNR of each experiment implies that as the number of steps increases, the error of guessing θ_1 has a reduced effect on the overall performance. We demonstrate the achievability of performance for any N-step design through the following theorem.

Theorem 4.1: For an N-step procedure, we need to design a sequence of input vectors $\{\mathbf{x}_i\}_{i=1}^{N}$ optimally under an average energy constraint to minimize the MSE in (4). Let $\mathcal{S} = \{\mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1}; \theta_1)\}_{i=1}^{N}$ be any design of the input parameters satisfying the following conditions:

- Average energy constraint:

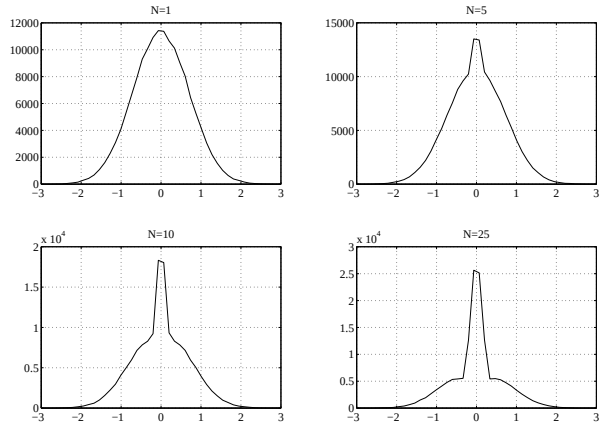


Fig. 5. Distribution of noise versus number of steps.

$\mathrm{E}\left[\sum_{i=1}^{N} \|\mathbf{x}_i\left(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1};\theta_1\right)\|^2\right] \le E_0$.

- Continuity - The design vector $\mathbf{x}_i\left(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1};\theta_1\right)$ is a continuous function of $\{\mathbf{y}_j\}_{j=1}^{i-1}$ or can assume the form of a thresholding function in (31).

Then there exists a θ_1-independent strategy whose performance can come arbitrarily close to $\mathrm{MSE}^{(N)}(\mathcal{S})$ which assumes the knowledge of parameter θ_1.

Proof: The proof is similar to the $N \times 2$-step strategy presented in Section IV, where the actual value of θ_1 in the optimal solution is replaced with a guess of θ_1. Refer to [42] for details. ■

V. Sequential Design for Vector Parameters

A general N-step procedure for the case of M unknown parameters can be written as

$$\mathbf{y}_i = \mathbf{H}(\mathbf{x}_i(\mathbf{y}_1,\ldots,\mathbf{y}_{i-1}))\boldsymbol{\theta} + \mathbf{n}_i, \quad i = 1,2,\ldots,N, \quad (34)$$

where $\boldsymbol{\theta}$ is an M-element vector, $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0},\mathbf{R}_\mathbf{n})$, and $\mathbf{H}(\mathbf{x})$ is a $K \times M$ matrix. For the multiple parameter case, MSE is no longer a scalar. Various criteria such as trace, minmax, determinant of the MSE matrix can be considered as measures of performance under the multiple unknown setting.

A. Trace Criteria

For the multiple parameter case, the MSE is a matrix and we consider the trace as a measure of performance, i.e., $\min_{\{\mathbf{x}_i\}_{i=1}^N} \mathrm{tr}(\mathrm{MSE}^{(N)}(\boldsymbol{\theta}))$. The problem of multiple parameter estimation is more complicated than estimation of a single parameter for the following reason. We showed in Section II-A that independent of the shape of $\mathbf{x}_i$, any non-adaptive energy allocation strategy is to assign all energy to the first step, i.e., a one-step strategy with energy E_0. But this is not true for the multiple parameter setting. Let us consider a simple example of estimating two parameters $\boldsymbol{\theta} = [\theta_1\ \theta_2]^T$ in the model $\mathbf{y} = \mathbf{H}(\mathbf{x})\boldsymbol{\theta} + \mathbf{n}$, where

$$\mathbf{H}(\mathbf{x}) = \begin{bmatrix} x_1 & x_2 \\ 0 & x_2 \end{bmatrix}, \quad (35)$$

$\mathbf{x} = [x_1\ x_2]^T$, $\mathbf{y} = [y_1\ y_2]^T$, $\mathbf{n} = [n_1\ n_2]^T \sim \mathcal{CN}(\mathbf{0},\mathbf{R}_\mathbf{n})$, and $\mathbf{R}_\mathbf{n} = \sigma^2\mathbf{I}$. Then for a one-step process, we have $\mathrm{MSE}^{(1)}(\theta_1) = 2\sigma^2/x_1^2$ and $\mathrm{MSE}^{(1)}(\theta_2) = \sigma^2/x_2^2$. Minimizing $\mathrm{tr}(\mathrm{MSE}^{(1)}(\boldsymbol{\theta})) = \mathrm{MSE}^{(1)}(\theta_1) + \mathrm{MSE}^{(1)}(\theta_2)$ over the energy constraint $\|\mathbf{x}\|^2 \le E_0 = 1$, we obtain $x_1 = x_2 = 1/\sqrt{2}$ and $\mathrm{tr}(\mathrm{MSE}^{(1)}_{\min}) = 6\sigma^2$. Now consider the following two-step non-adaptive strategy,

Step 1. $\mathbf{x} = [x_1\ 0]^T$, $y_1 = x_1\theta_1 + n_1$,
Step 2. $\mathbf{x} = [0\ x_2]^T$, $[1\ 1]\mathbf{y}_2 = 2x_2\theta_2 + [1\ 1]\mathbf{n}_2$.

Minimizing the $\mathrm{tr}(\mathrm{MSE}^{(2)}(\boldsymbol{\theta})) = \mathrm{MSE}^{(2)}(\theta_1) + \mathrm{MSE}^{(2)}(\theta_2) = \sigma^2/x_1^2 + \sigma^2/2x_2^2$ over the energy constraint, we obtain $x_1 = x_2 = 1/\sqrt{2}$ and $\mathrm{tr}(\mathrm{MSE}^{(2)}_{\min}) = 3\sigma^2$. This translates to a 3dB gain in SNR for the two-step non-adaptive strategy over the one-step approach. We control the shape of the input $\mathbf{x} = [x_1\ x_2]^T$ such that we have different energy allocation for each column of the matrix $\mathbf{H}$. By specifically designing the two-step non-adaptive strategy given in steps 1 and 2, we have reduced the estimation of the vector parameter $\boldsymbol{\theta} = [\theta_1,\theta_2]$ to two independent problems of estimating scalar parameters θ_1 and θ_2 respectively. For each of these scalar estimators, we design two N-step sequential procedures ($2N$ steps in total) as in Section IV-A for scalar controls x_1 and x_2 to obtain an improvement in performance of estimating $\boldsymbol{\theta}$. Applying the N-step design to both x_1 and x_2, we have $\mathrm{MSE}^{(N)}(\theta_1) = \mathcal{G}_N\mathrm{MSE}^{(2)}_{\min}(\theta_1)$ for the first N steps and $\mathrm{MSE}^{(N)}(\theta_2) = \mathcal{G}_N\mathrm{MSE}^{(2)}_{\min}(\theta_2)$ for the next N steps. Hence $\mathrm{tr}(\mathrm{MSE}^{(2N)}) = \mathcal{G}_N\mathrm{tr}(\mathrm{MSE}^{(2)}_{\min})$, where $\mathcal{G}_N$ is defined below (33). In other words, the MSE gains of the N-step procedure carry over to the vector parameter case as well.

B. Worst Case Error - Min Max Approach

The component wise MSE for estimating specific parameters is given by the diagonal elements of the matrix $\mathrm{MSE} = \mathrm{E}\left[(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^H\right]$. We seek to find the optimal energy allocation between the two design vectors, $\mathbf{x}_i(\{\mathbf{y}_j\}_{j=1}^{i-1}) = \mathbf{u}_\mathrm{m}\sqrt{E_0}\alpha_i(\{\mathbf{y}_j\}_{j=1}^{i-1})$, $i = 1,2$, that minimizes the worst case mean-squared error (*WC*-MSE) of the unknown parameters, where $\mathbf{u}_\mathrm{m}$ is any unit norm vector independent of past measurements, e.g., $\mathbf{u}_\mathrm{m}$ is chosen to minimize the one-step MSE. The ML estimate for a one-step process with energy E_0 and its corresponding MSE are given by

$$\widehat{\boldsymbol{\theta}}^{(1)} = \frac{1}{\sqrt{E_0}}\mathbf{W}_{\mathbf{u}_\mathrm{m}}\mathbf{H}(\mathbf{u}_\mathrm{m})^H\mathbf{R}_\mathbf{n}^{-1}\mathbf{y}_1, \ \mathrm{MSE}^{(1)} = \frac{1}{E_0}\mathbf{W}_{\mathbf{u}_\mathrm{m}}, \quad (36)$$

where

$$\mathbf{W}_{\mathbf{u}_\mathrm{m}} = (\mathbf{H}(\mathbf{u}_\mathrm{m})^H\mathbf{R}_\mathbf{n}^{-1}\mathbf{H}(\mathbf{u}_\mathrm{m}))^{-1}. \quad (37)$$

Define $\Phi(\mathbf{u},\mathrm{MSE}) = \mathbf{u}^H\mathrm{MSE}\mathbf{u}$. Then

$$\textit{WC}\text{-MSE} = \max_i \mathbf{e}_i^H\mathrm{MSE}\mathbf{e}_i = \max_i \Phi(\mathbf{e}_i,\mathrm{MSE}), \quad (38)$$

where $\mathbf{e}_i$ is an M-element vector with all zeros except for 1 in the i^{th} position. Then for a one-step process

$$\textit{WC}\text{-MSE}^{(1)} = \max_i \Phi(\mathbf{e}_i,\mathrm{MSE}^{(1)}) = \Phi(\mathbf{e}_{i^*},\mathrm{MSE}^{(1)}), \quad (39)$$

where i^* indicates the $\arg\max_i \Phi(\mathbf{e}_i, \text{MSE}^{(1)})$ and

$$\Phi(\mathbf{u}, \text{MSE}^{(1)}) = \frac{1}{E_0}\mathbf{u}^H \mathbf{W}_{\mathbf{u}_\text{m}} \mathbf{u}. \quad (40)$$

The set of observations for the two-step process are

$$\mathbf{y}_1 = \sqrt{E_0}\alpha_1 \mathbf{H}(\mathbf{u}_\text{m})\boldsymbol{\theta} + \mathbf{n}_1 \quad (41)$$

$$\mathbf{y}_2 = \sqrt{E_0}\alpha_2(\mathbf{y}_1)\mathbf{H}(\mathbf{u}_\text{m})\boldsymbol{\theta} + \mathbf{n}_2. \quad (42)$$

For a two-step procedure, we need to design α_1 and $\alpha_2(\mathbf{y}_1)$ to minimize *WC*-MSE$^{(2)}$. We show in [42] that

$$\Phi(\mathbf{u}, \text{MSE}^{(2)}) = \Phi(\mathbf{u}, \text{MSE}^{(1)}) \, \text{E}\left[\frac{\alpha_1^2 |\tilde{n}_1(\mathbf{y}_1;\boldsymbol{\theta})|^2 + \alpha_2^2(\tilde{n}_1(\mathbf{y}_1;\boldsymbol{\theta}))}{(\alpha_1^2 + \alpha_2^2(\tilde{n}_1(\mathbf{y}_1;\boldsymbol{\theta}))^2}\right], \quad (43)$$

where

$$\tilde{n}_1(\mathbf{y}_1;\boldsymbol{\theta}) = \frac{\mathbf{u}^H \mathbf{W}_{\mathbf{u}_\text{m}} \mathbf{H}(\mathbf{u}_\text{m})^H \mathbf{R}_\mathbf{n}^{-1}(\mathbf{y}_1 - \mathbf{H}(\mathbf{x}_1)\boldsymbol{\theta})}{\sqrt{\mathbf{u}^H \mathbf{W}_{\mathbf{u}_\text{m}} \mathbf{u}}} \quad (44)$$

is a $\mathcal{CN}(0,1)$ random variable. The error in (43) when minimized under the constraint $\alpha_1^2 + \text{E}\left[\alpha_2^2(\tilde{n}_1)\right] \leq 1$ is exactly the same minimization derived for the single parameter case in Section III. It follows that the optimal and suboptimal solutions to α_1 and $\alpha_2(\cdot)$ will hold for the multiple parameter case. In other words $\Phi(\mathbf{u}, \text{MSE}^{(2)}) \approx 0.6821\ \Phi(\mathbf{u}, \text{MSE}^{(1)})$. It follows that

$$\textit{WC}\text{-MSE}^{(2)} = \Phi(\mathbf{e}_{i^*}, \text{MSE}^{(2)}) \approx 0.6821\ \Phi(\mathbf{e}_{i^*}, \text{MSE}^{(1)}) = 0.6821\ \textit{WC}\text{-MSE}^{(1)} \quad (45)$$

and this performance can be achieved using a $\boldsymbol{\theta}$-independent strategy along similar lines to the derivation for the scalar parameter case in Section IV [42]. The reduction in MSE in (45) holds for any M, the number of unknown parameters, as i^*, the index of the worst case error, can always be computed from (39) and (40) for any $M \in \mathbb{N}$. A similar result can be derived for the N-step procedure.

VI. Applications of sequential estimation

A. MIMO Channel Estimation

It has been shown that multiple-input and multiple-output systems (MIMO) greatly increase the capacity of wireless systems [44]–[46] and hence MIMO has become an active area of research over the last decade [47], [48]. One important component in a MIMO system is the need to accurately estimate the channel state information (CSI) at the transmitter and receiver. This estimate has shown to play a crucial role in MIMO communications [49]. A recent and popular approach to channel estimation has been through the use of training sequences, i.e., known pilot signals are transmitted and channel is estimated using the received data and the pilot signals. A number of techniques for performing training based channel estimation have been proposed: maximum likelihood training method [50], least squares training [51], minimum mean squared estimation [52]. Recently, [40] proposed four different training methods for the flat block-fading MIMO system including the least squares and best linear unbiased estimator (BLUE) approach for the case of multiple LS channel estimates.

1) Problem Formulation: In order to estimate the $r \times t$ channel matrix $\boldsymbol{\Theta}$ for a MIMO system with t transmit and r receive antennas, $N \geq t$ training vectors $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_N]$ are transmitted. The corresponding set of received signals can be expressed as [40], [53]

$$\mathbf{R} = \boldsymbol{\Theta}\mathbf{X} + \mathbf{M}, \quad (46)$$

where $\mathbf{R} = [\mathbf{r}_1, \ldots, \mathbf{r}_N]$ is a $r \times N$ matrix, $\mathbf{M} = [\mathbf{m}_1, \ldots, \mathbf{m}_N]$ is the $r \times N$ matrix of sensor noise, $\mathbf{x}_i$ is the $t \times 1$ complex vector of transmitted signals, and $\mathbf{m}_i$ is the $r \times 1$ complex zero mean white noise vector. Let P_0 be the transmitted training power constraint, i.e., $\|\mathbf{X}\|_\text{F}^2 = P_0$, $\|\cdot\|_F$ indicates Frobenius norm ($\|\mathbf{X}\|_\text{F} = \sqrt{\text{tr}(\mathbf{X}^H\mathbf{X})}$) and σ^2 denote the variance of receiver noise. Though $\boldsymbol{\Theta}$ is random, we estimate $\boldsymbol{\Theta}$ for a particular realization corresponding to the block of received data. The task of channel estimation is to recover the channel matrix $\boldsymbol{\Theta}$ based on the knowledge of $\mathbf{X}$ and $\mathbf{R}$ as accurately as possible under a transmit power constraint on $\mathbf{X}$. The standard LS solution and the corresponding estimation error can then be written as

$$\widehat{\boldsymbol{\Theta}}_\text{LS} = \mathbf{R}\mathbf{X}^H(\mathbf{X}\mathbf{X}^H)^{-1}, \ \text{MSE}_\text{LS} = \frac{\sigma^2 t^2 r}{P_0}. \quad (47)$$

Assuming co-located transmitter and receiver arrays [54], [55] and multiple training periods available within the same coherency time (quasi-static) to estimate the channel, the set of received signals at the N time steps given by $\mathbf{R}_i = \boldsymbol{\Theta}\mathbf{X}_i + \mathbf{M}_i,\ i = 1, 2, \ldots, N$, can be rewritten in the following form:

$$\mathbf{y}_i = \mathbf{H}(\mathbf{X}_i)\boldsymbol{\theta} + \mathbf{n}_i, \quad i = 1, 2, \ldots, N, \quad (48)$$

where $\mathbf{y}_i = \text{vec}(\mathbf{R}_i), \boldsymbol{\theta} = \text{vec}(\boldsymbol{\Theta}), \mathbf{n}_i = \text{vec}(\mathbf{M}_i), \text{vec}(\cdot)$ denotes the column-wise concatenation of the matrix, and $\mathbf{H}(\mathbf{X}_i) = (\mathbf{X}_i \otimes \mathbf{I})^T$ is a linear function of the input $\mathbf{X}_i$, which is the same model described in (34). In [40], a method of linearly combining the estimates from each of the N stages was proposed and the MSE of the N stage estimator was shown to be $\text{MSE}_\text{LS}^{(N)} = \sigma^2 t^2 r / P_0$, where P_0 is the total power used in the N steps, i.e., $\sum_{i=1}^N \|\mathbf{X}_i\|_\text{F}^2 \leq P_0$. If there are enough training samples, we could completely control the matrix $\mathbf{H}(\mathbf{X}_i)$ through the input $\mathbf{X}_i$ and make $\mathbf{H}(\mathbf{X}_i)$ orthogonal. In this case (48) along with the average power constraint $\text{E}\left[\sum_i \|\mathbf{X}_i\|_\text{F}^2\right] \leq P_0$ can benefit from adaptive energy allocation designs in Sections IV-A and V-A, where the problem is then separable into rt independent estimation problems of scalar parameters. Having N steps in the training sequence suggests an N-step energy allocation strategy. Hence it follows that using our strategy we are guaranteed to achieve the optimal error given by $\text{MSE}^{(N)} \approx \mathcal{G}_N \sigma^2 t^2 r / P_0$, which we have shown to be at least 5dB (in 50 steps) better than any non-adaptive strategy.

B. Inverse Scattering Problem

The problem of imaging a medium using an array of transducers has been widely studied in many research areas such as mine detection, ultrasonic medical imaging [56],

foliage penetrating radar, non-destructive testing [57], and active audio. The goal in imaging is to detect and image small scatterers in a known background medium. A recent approach [58] uses the concept of time reversal, which works by exploiting the reciprocity of a physical channel, e.g., acoustic, optical, or radio-frequency. One implication of reciprocity is that a receiver can reflect back a time reversed signal, thereby focusing the signal at the transmitter source [59]. Furthermore, with suitable prefiltering and aperture, the signal energy can also be focused on an arbitrary spatial location. This analysis assumes the noiseless scenario. For the noisy case, maximum likelihood estimation of point scatterers was performed for both the single scattering and the multiple scattering models in [41]. We apply our concept of designing a sequence of measurements to image a medium of multiple scatterers using an array of transducers under a near-field approximation of the scatterers in the medium.

1) Problem Setting: We have N transducers located at positions $\{\mathbf{r}_k^a\}_{k=1}^N$, that transmit narrowband signals with center frequency ω rad/sec. The imaging area (or volume) is divided into V voxels at positions $\{\mathbf{r}_k^v\}_{k=1}^V$. The channel, denoted $\mathbf{a}_i$, between a candidate voxel i and the N transducers is given by the homogeneous Green's function as

$$\mathbf{a}_i = \left[\left(\frac{\exp(-j\omega/c\|\mathbf{r}_k^a - \mathbf{r}_i^v\|)}{\|\mathbf{r}_k^a - \mathbf{r}_i^v\|} \right)_{k=1\ldots N} \right]^T, \qquad (49)$$

where c is the speed of light and $j = \sqrt{-1}$. This channel model is a narrowband near-field approximation, which ignores the effect of multiple scattering and has been widely adopted in other scattering studies, e.g., [60]. Each voxel can be characterized by its scatter coefficient, e.g., radar cross-section (RCS), $\{\theta_v\}_{v=1}^V$, which indicates the proportion of the received field that is re-radiated. Thus the channel between the transmitted field and the measured backscattered field at the transducer array is $\mathbf{A}\text{diag}(\boldsymbol{\theta})\mathbf{A}^T$, where $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_V]$, $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_V]^T$, and $\text{diag}(\boldsymbol{\theta})$ denotes a $V \times V$ diagonal matrix with θ_i as its i^{th} diagonal element.

The probing mechanism for imaging of the scatter cross-section follows a sequential process, generating the following sequence of noise contaminated signals,

$$\begin{aligned} \mathbf{y}_i &= \mathbf{A}\text{diag}(\boldsymbol{\theta})\mathbf{A}^T\mathbf{x}_i + \mathbf{n}_i \\ &= \mathbf{H}(\mathbf{x}_i)\boldsymbol{\theta} + \mathbf{n}_i, \ i = 1, 2, \ldots, N, \end{aligned} \qquad (50)$$

where $\mathbf{H}(\mathbf{x}_i) = \mathbf{A}\text{diag}(\mathbf{A}^T\mathbf{x}_i)$. The noises $\{\mathbf{n}_i\}$ are i.i.d $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$ random vectors. The goal is to find estimates for the scattering coefficients $\boldsymbol{\theta}$ under the average energy constraint to minimize the MSE. If $\mathbf{A}$ is a square matrix, then we can condition $\text{diag}(\mathbf{A}^T\mathbf{x}_i)$ to have a single non zero component on any one of the diagonal elements, which translates to isolating the i^{th} column of $\mathbf{H}$ for any i. As in Section V-A, we can perform V independent N-step experiments to guarantee the N-step gains of at least 5dB over the standard single step ML estimation for imaging [41]. If we are interested in optimally estimating any linear combination of the scattering coefficients, then the sequential strategy proposed in Section V-B can be used to achieve improvement in performance.

VII. Conclusions

In this paper we considered the N-step adaptive waveform amplitude design problem for estimating parameters of an unknown channel under average energy constraints. For a two-step problem, we found the optimal energy allocation at the second step as a function of the first measurement for a scalar parameter in the linear Gaussian model. We showed that this two-step adaptive strategy resulted in an improvement of at least 1.65dB over the optimal non-adaptive strategy. We then designed a suboptimal N-stage energy allocation procedure based on the two-step approach and demonstrated gains of more than 5dB in $N = 50$ steps. We extended our results to the case of vector parameters and provided applications of our design to MIMO and inverse scattering channel models.

References

[1] F. Daum, "A system approach to multiple target tracking," *Multitarget-Multisensor Tracking:Application and Advances*, vol. 2, pp. 149–181, 1992.

[2] M. Athans and F. C. Schweppe, "Optimal waveform design via control theoretic principles," *Inform. Control*, vol. 10, pp. 335–377, 1967.

[3] F. C. Schweppe and D. L. Gray, "Radar signal design subject to simultaneous peak and average constraints," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 13–26, 1966.

[4] D. J. Kershaw and R. J. Evans, "Optimal waveform selection for tracking systems," *IEEE Trans. Inform. Theory*, vol. 40, no. 5, pp. 1536–1550, 1994.

[5] ——, "Waveform selective probabilistic data association," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 4, pp. 1180–1188, 1997.

[6] S. P. Sira, A. P. Suppappola, and D. Morrell, "Waveform scheduling in wideband environments," *IEEE Intl. Conf. Acoust. Speech and Signal Proc.*, 2006.

[7] J. S. Evans and V. Krishanmurthy, "Optimal sensor scheduling for hidden markov model state estimation," *Int. J. Contr.*, vol. 74, no. 18, pp. 1737–1742, 2001.

[8] M. Athans, "On the determination of optimal costly measurement strategies for linear stochastic systems," *Automatica*, vol. 8, pp. 397–412, 1972.

[9] L. Meier, J. Perschon, and R. M. Dressier, "Optimal control of measurement subsystems," *IEEE Trans. Automat. Contr.*, vol. AC-12, pp. 528–536, 1967.

[10] V. Krishanmurthy, "Algorithms for optimal scheduling and management of hidden markov model sensors," *IEEE Trans. Signal Processing*, vol. 50, no. 6, pp. 1382–1396, 2002.

[11] J. B. S. Haldane, "On a method of estimating frequencies," *Biometrika*, vol. 33, pp. 222–225, 1945.

[12] V. Z. Borisov and V. V. Konev, "On sequential parameter estimation in discrete time process," *Automat. Remote Contr.*, vol. 38, pp. 58–64, 1977.

[13] P. R. Kumar and P. Varaiya, *Stochastic Systems - Estimation, Identificiation and Adaptive Control*. Englewood Cliffs, N.J: Prentice-Hall, 1986.

[14] C. M. Stein, "Inadmissibility of the usual estimator for the mean of a multivariate normal distribution," *Proc. 3rd Berkeley Symp. Math. Stat. Prob.*, vol. 1, pp. 197–206, 1956.

[15] W. James and C. M. Stein, "Estimation with quadratic loss," *Proc. 4th Berkeley Symp. Math. Stat. Prob.*, vol. 1, pp. 361–379, 1961.

[16] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York, N.Y: Springer-Verlag, 1988.

[17] D. Seigmund, "Herbert Robbins and sequential analysis," *Ann. Statist.*, vol. 31, pp. 349–365, 2003.

[18] H. Chernoff, "Sequential design of experiments," *Ann. Math. Statist.*, vol. 30, pp. 755–770, 1959.

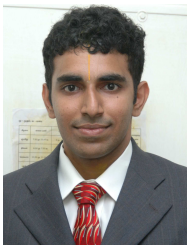
[19] D. Richter, "Two-stage experiments for estimating a common mean," *Ann. Math. Statist.*, vol. 31, pp. 1164–1173, 1960.

[20] S. G. Ghurye and H. Robbins, "Two-stage experiments for estimating the difference between means," *Biometrika*, vol. 41, pp. 146–152, 1954.

[21] J. K. B. Gray, "Sequential selection of experiments," *Ann. Math. Statist.*, vol. 39, pp. 1953–1977, 1968.

[22] J. Kiefer, "Optimal experimental designs," *J. Roy. Statist. Soc. Ser. B(Methodological)*, vol. 21, pp. 272–319, 1959.

[23] C. Stein, "A two sample test for a linear hypothesis whose power is independent of the variance." *Ann. Math. Statist.*, vol. 16, pp. 243–258, 1945.
[24] J. B. S. Haldane, "On a method of estimating frequencies," *Biometrika*, vol. 33, pp. 222–225, 1945.
[25] E. S. Page, "Continuous inspection schemes," *Biometrika*, vol. 41, pp. 100–114, 1954.
[26] H. E. Robbins, "Some aspects of sequential design of experiments." *Bull. Amer. Math. Soc.*, vol. 58, pp. 527–535, 1952.
[27] W. A. Shewhart, *Economic Control of Manufactured Products*. Newyork: Van Nostrand Reinhold, 1931.
[28] A. Wald, "Sequential tests of statistical hypotheses," *Ann. Math. Statist.*, vol. 16, pp. 117–186, 1945.
[29] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, pp. 400–407, 1951.
[30] V. V. Konev and T. L. Lai, "Estimators with prescribed precision in stochastic regression models," *Sequential Anal.*, vol. 14, pp. 179–192, 1995.
[31] F. J. Anscombe, "Sequential estimation." *J. Roy. Statist. Soc. Ser. B*, vol. 15, pp. 1–29, 1953.
[32] C. Stein, "Some problems in sequential estimation." *Econometrica*, vol. 17, pp. 77–78, 1949.
[33] L. S. Mayer and T. A. Willke, "On biased estimation in linear models," *Technometr.*, vol. 15, pp. 497–508, 1973.
[34] A. N. Tikhonov and V. Y. Arsenin, *Solution to Ill-Posed Problems*. Washington D.C: V. H. Winston, 1977.
[35] Y. C. Eldar and A. V. Oppenheim, "Covariance shaping least-squares estimation," *IEEE Trans. Signal Processing*, vol. 51, no. 3, pp. 686–697, 2003.
[36] V. V. Fedorov, *Theory of Optimal Experiments.* New York: Academic Press, 1972.
[37] M. B. Zarrop, *Optimal Experiment Design for Dynamic System Identification.* New York: Springer-Verlag, 1979.
[38] B. Baygun and A. O. Hero III, "Optimal simultaneous detection and estimation under a false alarm constraint," *IEEE Trans. on Info. Thy.*, vol. 41, no. 3, pp. 688–703, 1995.
[39] M. Mobed and W. L. Root, "Estimation of constrained parameters in a linear model with multiplicative and additive noise," *IEEE Trans. on Info. Thy.*, vol. 40, no. 1, pp. 56–66, 1994.
[40] M. Biguesh and A. B. Gershman, "Training based mimo channel estimation: A study of estimator tradeoffs and optimal trainign signals," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 884–892, 2006.
[41] G. Shi and A. Nehorai, "Maximum likelihood estimation of point scatterers for computational time-reversal imaging," *Communications in Information and Systems*, vol. 5, no. 2, pp. 227–256, 2005.
[42] R. Rangarajan, R. Raich, and A. O. Hero III, "Adaptive energy allocation for estimation in linear models," *Technical Report #378, Communications and Signal Processing Laboratory, University of Michigan, Ann Arbor*, Feb. 2007.
[43] ——, "Optimal experimental design for an inverse scattering problem," *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Proc.*, vol. 4, pp. 1117–1120, 2005.
[44] I. Telatar, "Capacity of multiple antenna gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–95, 1999.
[45] C. C. Martin, J. H. Winters, and N. R. Sollenberger, "Multiple-input multiple-output (mimo) radio channel measurements," *Proc. VTC'00*, vol. 2, pp. 774–779, 2000.
[46] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Commun.*, vol. 6, pp. 311–335, 1998.
[47] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1451–1458, 1998.
[48] F. Rashid-Farrokhi, K. J. R. Liu, and L. Tassiulas, "Transmit beamforming and power control for cellular wireless systems," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1437–1450, 1998.
[49] C. Buianu and L. Tong, "Channel estimation for space-time orthogonal block codes," *IEEE Trans. Signal Process.*, vol. 50, pp. 2515–2528, 2002.
[50] T. L. Marzetta, "Blast training: Estimating channel characteristics for high capacity space-time wireless," *Proc. 37th Annual Allerton Conf. Communications, Control, Computing*, 1999.
[51] Y. Li, "Optimal training sequences for ofdm systems with multiple transmite antennas," *Proc. GLOBECOM*, vol. 3, pp. 1478–1482, 2000.
[52] A. Scaglione and A. Vosoughi, "Turbo estimation of channel and symbols in precoded mimo systems," *Proc. ICASSP*, vol. 4, pp. 413–416, 2004.
[53] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireles Communications*. Cambridge, U.K: Cambridge Univ. Press, 2003.
[54] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, 1998.
[55] O. Simeone and U. Spagnolini, "Channel estimation for block-fading frequency selective rayleigh mimo channels," *IEEE Trans. Signal Process.*, vol. 52, no. 11, pp. 3265–3277, 2004.
[56] M. Fink and C. Pradha, "Acoustic time-reversal mirrors," *Inverse Problems*, vol. 17, pp. 1–38, 2001.
[57] E. Kerbrat, C. Prada, and M. Fink, "Ultrasonic nondestructive testing of scattering media using the decomposition of the time-reversal operator," *IEEE Trans. on Ultra., Ferro., and Freq., Control*, vol. 49, pp. 1103–1112, 2002.
[58] L. Borcea, G. C. Papanicolaou, C. Tsogka, and J. Berryman, "Imaging and time reversal in random media," *Inverse Problems*, vol. 18, pp. 1247–1279, 2002.
[59] M. Fink, "Time reversal acoustics," *Physics Today*, pp. 34–40, 1997.
[60] D. Bliss, K. W. Forysthe, A. O. Hero III, and A. F. Yegulalp, "Environmental issues for mimo capacity," *Signal Processing*, vol. 50, no. 9, pp. 2128–2142, 2002.



Raghuram Rangarajan was born in India in 1981. He received the B.Tech. degree from the Department of Electrical Engineering, Indian Institute of Technology, Madras, in 2002 and the M.S. degree in Electrical Engineering:Systems from the University of Michigan, Ann Arbor, in 2004. He is currently pursuing the Ph.D. degree in the Department of Electrical Engineering and Computer Science at the University of Michigan.

His research interests include statistical signal processing and design of experiments for sensor networks and agile sensing applications.

Raviv Raich received the B.Sc. and M.Sc. degrees in electrical engineering from Tel-Aviv University, Tel-Aviv, Israel, in 1994 and 1998, respectively and the Ph.D. degree in electrical engineering from Georgia Institute of Technology, Atlanta, Georgia, in 2004.

Between 1999 and 2000, he worked as a researcher with the communications team, Industrial Research Ltd., Wellington, New Zealand. Currently, he is a postdoctoral fellow at the university of Michigan, Ann Arbor, Michigan. His main research interest is in statistical signal processing with specific focus on manifold learning, sparse signal reconstruction, adaptive sensing. Other research interests lie in the area of statistical signal processing for communications, estimation and detection theory.

Alfred O. Hero III received the B.S. (summa cum laude) from Boston University (1980) and the Ph.D from Princeton University (1984), both in Electrical Engineering. Since 1984 he has been with the University of Michigan, Ann Arbor, where he is a Professor in the Department of Electrical Engineering and Computer Science and, by courtesy, in the Department of Biomedical Engineering and the Department of Statistics. His recent research interests have been in areas including: inference in sensor networks, adaptive sensing, bioinformatics, inverse problems, and statistical signal and image processing. He is a Fellow of the Institute of Electrical and Electronics Engineers (IEEE) and has received a IEEE Signal Processing Society Meritorious Service Award (1998), a IEEE Signal Processing Society Best Paper Award (1998), and the IEEE Third Millenium Medal (2000). Alfred Hero is currently President of the IEEE Signal Processing Society (2006-2008).