Adaptive Energy Allocation for Estimation in Linear Models

Raghuram Rangarajan, Raviv Raich, Alfred O. Hero III

February 27, 2007

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.

1 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.

1.1 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]. In table 1, we compare our work with existing literature via different categories.

1.2 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.

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 1: 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.

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 chapter, we formulate a problem of sequentially selecting waveform amplitudes for estimating deterministic parameters of a linear Gaussian channel model under an average energy constraint over the waveforms and over the number of transmissions. In Section 2, 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 3, closed-form expressions for the optimal design parameters (e.g., energy allocation to the waveforms in the adaptive sensing context) and the corresponding minimum MSE in the single parameter (e.g., scatter coefficients in imaging, channel coefficients in channel estimation) case are derived for a two-step procedure (two time steps). In Section 4, we provide a suboptimal design for the two-step strategy, which takes into consideration a peak power constraint and achieves near optimal performance. Since the optimal solution requires the knowledge of parameters to be estimated, it is shown in Section 6 that the performance of this omniscient solution can be achieved with a parameter independent strategy. In Section 7, 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 8. Finally in Section 9, we show the applicability of this framework by recasting the problems of channel estimation and radar imaging to fit the statistical model of the sequential parameter estimation problem and applying the results from the previous sections to show the advantages of our approach over current literature for practical applications.

2 Problem Statement

We denote vectors in $\mathbb{C}^M$ by boldface lower case letters and matrices in $\mathbb{C}^{M\times N}$ by boldface uppercase letters. The symbol $\|\cdot\|$ refers to the l_2-norm of a vector, i.e., $\|\mathbf{x}\| = \sqrt{\mathbf{x}^H\mathbf{x}}$, where $(\cdot)^H$ denotes the conjugate transpose. The terms MSE and SNR are abbreviations to mean-squared error and signal-to-noise ratio, respectively. Let $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_M]$ 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 circularly symmetric complex Gaussian random variable with zero mean and covariance $\sigma^2\mathbf{I}$ denoted by $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$. 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 parameter $\boldsymbol{\theta}$ can then be written as

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

For the case of a scalar parameter θ_1, the observations are

$$\mathbf{y}_i = \mathbf{h}_1(\mathbf{x}_i)\theta_1 + \mathbf{n}_i, \quad i = 1, 2, \ldots, N. \tag{3}$$

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, $\mathrm{E}\left[\sum_{i=1}^N \|\mathbf{x}_i\|^2\right] \leq E_0$, where E_0 is the total available energy and $\mathrm{E}[\cdot]$ denotes the statistical expectation. The ML estimator of θ_1 for the N-step procedure is given by

$$\hat{\theta}_1^{(N)} = \frac{\sum_{i=1}^N \mathbf{h}_1(\mathbf{x}_i)^H \mathbf{y}_i}{\sum_{i=1}^N \|\mathbf{h}_1(\mathbf{x}_i)\|^2} \tag{4}$$

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

$$\mathrm{MSE}^{(N)}\left(\{\mathbf{x}_i\}_{i=1}^N\right) = \mathrm{E}\left[\left|\frac{\sum_{i=1}^N \mathbf{h}_1(\mathbf{x}_i)^H \mathbf{n}_i}{\sum_{i=1}^N \|\mathbf{h}_1(\mathbf{x}_i)\|^2}\right|^2\right]. \tag{5}$$

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,

$$\mathcal{E}\left[\{\mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})\}_{i=1}^N\right] = \mathrm{E}\left[\sum_{i=1}^N \|\mathbf{x}_i\|^2\right]. \tag{6}$$

The average energy constraint can be written as

$$\mathcal{E}\left[\{\mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})\}_{i=1}^N\right] = \mathrm{E}\left[\sum_{i=1}^N E_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})\right] \leq E_0. \tag{7}$$

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 (5) under the average energy constraint in (7).

2.1 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 $\sum_{i=1}^N E_i \leq E_0$. Simplifying the expression for MSE in (5), we have

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

Using the fact that $\{\mathbf{n}_i\|_{i=1}^N$ are i.i.d $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$, we obtain

$$\begin{aligned}
\mathrm{MSE}^{(N)} &= \frac{\sigma^2}{\sum_{i=1}^N \|\mathbf{h}_1(\mathbf{x}_i)\|^2} \\
&= \frac{\sigma^2}{\sum_{i=1}^N E_i \frac{\|\mathbf{h}_1(\mathbf{x}_i)\|^2}{\|\mathbf{x}_i\|^2}} \\
&\geq \frac{\sigma^2}{E_0 \lambda_{\mathrm{m}}(\boldsymbol{H}_1)},
\end{aligned} \tag{8}$$

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

$$\lambda_{\mathrm{m}}(\boldsymbol{H}_1) = \max_{\mathbf{x}} (\mathbf{x}^H \boldsymbol{H}_1^H \boldsymbol{H}_1 \mathbf{x})/(\mathbf{x}^H \mathbf{x}) = \max_{\mathbf{x}} \|\mathbf{h}_1(\mathbf{x})\|^2 / \|\mathbf{x}\|^2. \tag{9}$$

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

$$\mathrm{SNR}^{(N)} = \frac{\lambda_{\mathrm{m}}(\boldsymbol{H}_1)\mathcal{E}\left[\{\mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1})\}_{i=1}^N\right]}{\sigma^2}. \tag{10}$$

Then the average energy constraint in (7) is equivalent to $\mathrm{SNR}^{(N)} \leq \mathrm{SNR}_0$, where $\mathrm{SNR}_0 = \lambda_{\mathrm{m}}(\boldsymbol{H}_1)E_0/\sigma^2$. We show in Appendix 11 that the problem of minimizing $\mathrm{MSE}^{(N)}$ subject to $\mathrm{SNR}^{(N)} \leq \mathrm{SNR}_0$ is equivalent to minimizing $\mathrm{MSE}^{(N)} \times \mathrm{SNR}^{(N)}$. Thus we use the two minimization criteria interchangeably in the remainder of this paper. The product of MSE and SNR is

$$\mathrm{MSE}^{(N)} \times \mathrm{SNR}^{(N)} = \mathrm{E}\left[\left|\frac{\sum_{i=1}^N \mathbf{h}_1(\mathbf{x}_i)^H \mathbf{n}_i}{\sum_{i=1}^N \|\mathbf{h}_1(\mathbf{x}_i)\|^2}\right|^2\right] \frac{\lambda_{\mathrm{m}}(\boldsymbol{H}_1)\mathrm{E}\left[\sum_{i=1}^N \|\mathbf{x}_i\|^2\right]}{\sigma^2} \tag{11}$$

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

$$\mathrm{MSE}_{\mathrm{min}}^{(1)} \times \mathrm{SNR}_0 = 1. \tag{12}$$

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 $\mathrm{MSE}^{(N)} \times \mathrm{SNR}_0 < 1$ is also of interest. We first look at a two-step sequential design procedure. A word of caution: in Sections 3 and 4 we develop optimal and suboptimal strategies where the solutions require the knowledge of the unknown parameter θ_1. However, in Section 6 we present a θ_1-independent design which asymptotically achieves the performance of the 'omniscient' strategies.

3 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

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{h}_1(\mathbf{x}_1)\theta_1 + \mathbf{n}_1 \quad (13) \\ \mathbf{y}_2 &= \mathbf{h}_1(\mathbf{x}_2(\mathbf{y}_1))\theta_1 + \mathbf{n}_2. \quad (14) \end{aligned}$$

The ML estimator of θ_1 for a two-step procedure from (4) is

$$\hat{\theta}_1^{(2)} = \frac{\mathbf{h}_1(\mathbf{x}_1)^H \mathbf{y}_1 + \mathbf{h}_1(\mathbf{x}_2)^H \mathbf{y}_2}{\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2} \quad (15)$$

and its MSE from (5) is given by

$$\mathrm{MSE}^{(2)}(\mathbf{x}_1, \mathbf{x}_2) = \mathrm{E}\left[\frac{|\mathbf{h}_1(\mathbf{x}_1)^H \mathbf{n}_1 + \mathbf{h}_1(\mathbf{x}_2)^H \mathbf{n}_2|^2}{(\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2)^2}\right]. \quad (16)$$

We assume that the shape of the optimal designs, i.e., $\{\mathbf{x}_i/\|\mathbf{x}_i\|\}$ is the one-step optimum given by $\mathbf{v}_\mathrm{m}$ defined below (8) 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

$$\begin{aligned} \mathbf{x}_1 &= \sqrt{E_0}\ \alpha_1 \mathbf{v}_\mathrm{m} \quad (17) \\ \mathbf{x}_2(\mathbf{y}_1) &= \sqrt{E_0}\ \alpha_2(\mathbf{y}_1)\mathbf{v}_\mathrm{m}, \quad (18) \end{aligned}$$

where α_1 and $\alpha_2(\cdot)$ are real-valued scalars. The average energy constraint from (7) can then be written as

$$\mathrm{E}\left[\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)\right] \leq 1. \quad (19)$$

We use Lagrangian multipliers to minimize the MSE in (16) with respect to α_1 and $\alpha_2(\cdot)$ under the energy constraint in (19). Substituting for $\mathbf{x}_1$ and $\mathbf{x}_2(\mathbf{y}_1)$ given by (17) and (18) respectively in (16) and adding the Lagrangian constraint we obtain the objective function to be minimized as

$$\begin{aligned} & \mathrm{MSE}^{(2)}(\mathbf{x}_1, \mathbf{x}_2) + \gamma(\mathrm{E}\left[\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)\right]) \\ = \ & \mathrm{E}\left[\frac{|\mathbf{h}_1(\mathbf{x}_1)^H \mathbf{n}_1 + \mathbf{h}_1(\mathbf{x}_2)^H \mathbf{n}_2|^2}{(\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2)^2}\right] + \gamma\left(\alpha_1^2 + \mathrm{E}\left[\alpha_2^2(\mathbf{y}_1)\right]\right) \end{aligned}$$

Using linearity of $\mathbf{h}_1(\cdot)$, the objective function can be written as

$$\frac{1}{E_0}\mathrm{E}\left[\frac{|\alpha_1 \mathbf{h}_1(\mathbf{v}_\mathrm{m})^H \mathbf{n}_1 + \alpha_2(\mathbf{y}_1)\mathbf{h}_1(\mathbf{v}_\mathrm{m})^H \mathbf{n}_2|^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2 \|\mathbf{h}_1(\mathbf{v}_\mathrm{m})\|^4} + \gamma E_0\left(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)\right)\right]$$

Taking the expectation over $\mathbf{n}_2$, the objective function becomes

$$\frac{1}{E_0}\mathrm{E}\left[\frac{\alpha_1^2|\mathbf{h}_1(\mathbf{v}_\mathrm{m})^H\mathbf{n}_1|^2+\alpha_2^2(\mathbf{y}_1)\|\mathbf{h}_1(\mathbf{v}_\mathrm{m})\|^2\sigma^2}{(\alpha_1^2+\alpha_2^2(\mathbf{y}_1))^2\|\mathbf{h}_1(\mathbf{v}_\mathrm{m})\|^4}+\gamma E_0\left(\alpha_1^2+\alpha_2^2(\mathbf{y}_1)\right)\right]$$

Dividing numerator and denominator by $\|\mathbf{h}_1(\mathbf{v}_\mathrm{m})\|^2\sigma^2$, the minimization criterion simplifies to

$$\begin{aligned}
&= \frac{1}{\mathrm{SNR}_0}\mathrm{E}\left[\frac{\alpha_1^2\left|\tilde{n}_1(\mathbf{y}_1;\theta_1)\right|^2+\alpha_2^2(\mathbf{y}_1)}{(\alpha_1^2+\alpha_2^2(\mathbf{y}_1))^2}+\gamma\mathrm{SNR}_0\left(\alpha_1^2+\alpha_2^2(\mathbf{y}_1)\right)\right] \\
&= \frac{1}{\mathrm{SNR}_0}\frac{1}{\alpha_1^2}\mathrm{E}\left[\frac{1}{\left(1+\frac{\alpha_2^2(\mathbf{y}_1)}{\alpha_1^2}\right)}-\frac{1-\left|\tilde{n}_1(\mathbf{y}_1;\theta_1)\right|^2}{\left(1+\frac{\alpha_2^2(\mathbf{y}_1)}{\alpha_1^2}\right)^2}+\gamma^{'}\left(1+\frac{\alpha_2^2(\mathbf{y}_1)}{\alpha_1^2}\right)\right] \quad (20)
\end{aligned}$$

where $\tilde{n}_1(\mathbf{y}_1;\theta_1)=\frac{\mathbf{h}_1(\mathbf{v}_\mathrm{m})^H}{\|\mathbf{h}_1(\mathbf{v}_\mathrm{m})\|}\left(\frac{\mathbf{y}_1-\mathbf{h}_1(\mathbf{x}_1)\theta_1}{\sigma}\right)=\frac{\mathbf{h}_1(\mathbf{v}_\mathrm{m})^H}{\|\mathbf{h}_1(\mathbf{v}_\mathrm{m})\|}\frac{\mathbf{n}_1}{\sigma}$ is a zero mean unit variance complex Gaussian random variable and $\gamma^{'}=\gamma\alpha_1^4\mathrm{SNR}_0$. Since the optimal solution to $\alpha_2(\mathbf{y}_1)$ depends on $\mathbf{y}_1$ only through the function $\tilde{n}_1(\mathbf{y}_1;\theta_1)$, we denote the solution as $\alpha_2\left(\tilde{n}_1(\mathbf{y}_1;\theta_1)\right)$. Let $g\left(\tilde{n}_1(\mathbf{y}_1;\theta_1)\right)=\left(1+\frac{\alpha_2^2(\tilde{n}_1(\mathbf{y}_1;\theta_1))}{\alpha_1^2}\right)$. Then the objective function can be written as

$$\frac{1}{\mathrm{SNR}_0}\frac{1}{\alpha_1^2}\mathrm{E}\left[\frac{1}{g\left(\tilde{n}_1(\mathbf{y}_1;\theta_1)\right)}-\frac{1-\left|\tilde{n}_1(\mathbf{y}_1;\theta_1)\right|^2}{g^2\left(\tilde{n}_1(\mathbf{y}_1;\theta_1)\right)}+\gamma^{'}g\left(\tilde{n}_1(\mathbf{y}_1;\theta_1)\right)\right] \quad (21)$$

Differentiating and setting the objective function to zero, we have

$$g^3-\frac{1}{\gamma^{'}}g+2\frac{1-|\tilde{n}_1|^2}{\gamma^{'}}=0. \quad (22)$$

The function g that minimizes MSE is the root of the third-order polynomial in (22), 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 $\gamma^{'}$ is denoted by $g_{\gamma^{'}}(\tilde{n}_1)$. Also $\mathrm{E}\left[g_{\gamma^{'}}(\tilde{n}_1)\right]=\frac{1}{\alpha_1^2}$. Therefore, finding α_1 that minimizes MSE is equivalent to finding $\gamma^{'}$ that minimizes MSE. We obtain $g_{\gamma^{'}}(\tilde{n}_1)$ for every $\gamma^{'}$ and use a brute force grid search to find the optimal $\gamma^{'}$ that minimizes the expression in (21). 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

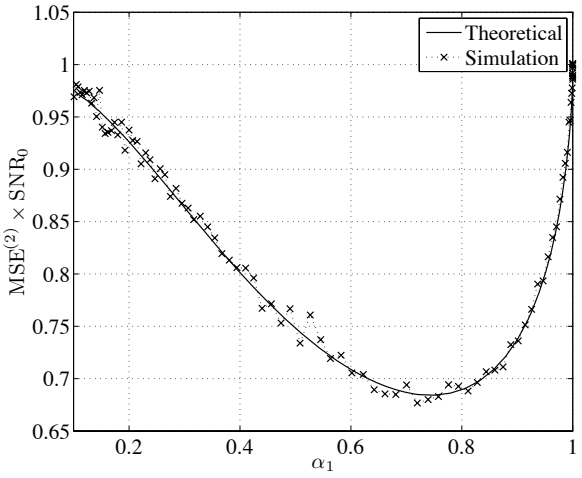


Figure 1: Reduction in MSE for varying values of α_1.

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

$$\mathrm{MSE}^{(2)}_{\min} \times \mathrm{SNR}_0 \quad \approx \quad 0.68. \tag{23}$$

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 $\mathrm{MSE}^{(1)}_{\min} \times \mathrm{SNR}_0 = 1$. $\mathrm{MSE}^{(2)} \times \mathrm{SNR}_0$ is plotted for various values of α_1 using both simulations (dotted) and theoretically (solid) in Fig. 1.

The theoretical performance curve (solid) was generated by evaluating the $\mathrm{MSE}^{(2)}$ in (21) (without the constraint term) for various values of $\gamma^{'}$ (or α_1). Since the expectation in (21) depends only on random variable $\tilde{n}_1$, we construct a fine grid of $\tilde{n}_1$ and approximate the integral induced by the expectation as a Riemann sum. For the simulation curve (dotted), we generate 10000 samples of $\mathbf{n}_1$ and $\mathbf{n}_2$ distributed as $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$. Using $\mathbf{n}_1$ and $\mathbf{x}_1$ from (17), we generate 10000 samples of $\mathbf{y}_1$ from (13). Using samples of $\mathbf{y}_1$, we generate samples of $\mathbf{x}_2(\mathbf{y}_1)$ from (18) and obtain 10000 samples of $\mathbf{y}_2$ from (14) using $\mathbf{x}_2(\mathbf{y}_1)$ and $\mathbf{n}_2$. We obtain an estimate of $\mathrm{MSE}^{(2)}$ by evaluating the expectation in (16) through numerical averaging over the realizations.

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. 2 (solid) is a thresholding function, i.e., α_2^* is zero for $|\tilde{n}_1|^2 \leq$

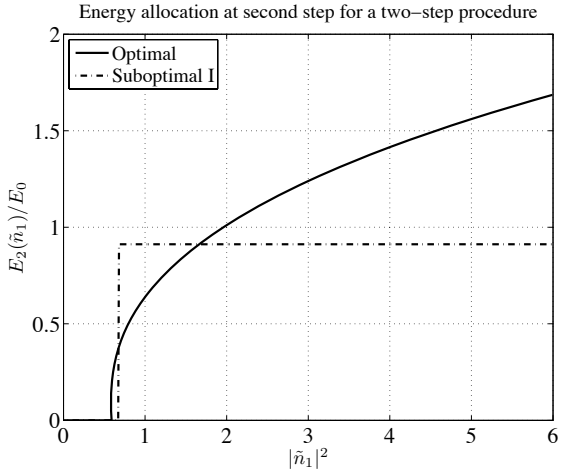


Figure 2: Plot of the optimal and suboptimal solution to the normalized energy transmitted at the second stage as functions of received signal at first stage.

0.59. This solution implies that when the actual realization of the normalized noise along $\mathbf{h}_1(\mathbf{v}_\mathrm{m})$ in the first step is small enough, then the second measurement becomes unnecessary. On the other hand, when the normalized noise along $\mathbf{h}_1(\mathbf{v}_\mathrm{m})$ 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 be possible. Hence, in the following section we look at a suboptimal solution which takes into account this constraint and still achieves near optimal performance.

4 Omniscient Suboptimal Two-step Strategy

The optimal solution in Section 3 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

$$\mathbf{x}_1 = \mathbf{v}_{\mathrm{m}} \sqrt{E_0}\, \alpha_1 \tag{24}$$

$$\mathbf{x}_2 = \mathbf{v}_{\mathrm{m}} \sqrt{E_0}\, \alpha_2 I\left(\left|\frac{\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\|}\frac{\mathbf{n}_1}{\sigma}\right|^2 > \rho\right) = \mathbf{v}_{\mathrm{m}} \sqrt{E_0}\, \alpha_2\, I\left(|\tilde{n}_1|^2 > \rho\right) \tag{25}$$

where $\tilde{n}_1$ is defined below (20), 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

$$\mathrm{SNR}^{(2)} = \mathrm{SNR}_0 \left(\alpha_1^2 + \alpha_2^2 P\left(|\tilde{n}_1|^2 > \rho\right)\right). \tag{26}$$

The MSE of the ML estimator under this suboptimal solution using (16) is

$$\begin{aligned} \mathrm{MSE}^{(2)} &= \mathrm{E}\left[\frac{|\mathbf{h}_1(\mathbf{x}_1)^H\mathbf{n}_1 + \mathbf{h}_1(\mathbf{x}_2)^H\mathbf{n}_2|^2}{(\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2)^2}\right] \\ &= \frac{1}{\mathrm{SNR}_0}\mathrm{E}\left[\frac{\alpha_1^2|\tilde{n}_1|^2 + \alpha_2^2}{\left(\alpha_1^2+\alpha_2^2\right)^2} I\left(|\tilde{n}_1|^2 \geq \rho\right)\right] + \frac{1}{\mathrm{SNR}_0}\mathrm{E}\left[\frac{|\tilde{n}_1|^2}{\alpha_1^2} I\left(|\tilde{n}_1|^2 < \rho\right)\right]. \end{aligned} \tag{27}$$

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 (27) and (26), we obtain

$$\begin{aligned} \mathrm{MSE}^{(2)} &= \frac{1}{\mathrm{SNR}_0}\frac{1}{\left(\alpha_1^2+\alpha_2^2\right)}\left(\mathrm{E}\left[(\beta|\tilde{n}_1|^2 + (1-\beta))I(|\tilde{n}_1|^2 \geq \rho) + \frac{|\tilde{n}_1|^2}{\beta} I(|\tilde{n}_1|^2 < \rho)\right]\right), \\ \mathrm{SNR}^{(2)} &= \mathrm{SNR}_0\left(\alpha_1^2+\alpha_2^2\right)\left(\beta + (1-\beta)P(|\tilde{n}_1|^2 \geq \rho)\right). \end{aligned}$$

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 expressions for $\mathrm{MSE}^{(2)}$ and $\mathrm{SNR}^{(2)}$ simplify to

$$\begin{aligned} \mathrm{MSE}^{(2)} &= \frac{1}{\mathrm{SNR}_0}\frac{1}{\left(\alpha_1^2+\alpha_2^2\right)}\left(\beta\rho e^{-\rho} + e^{-\rho} + \frac{1}{\beta}(1-(1+\rho)e^{-\rho})\right), \\ \mathrm{SNR}^{(2)} &= \mathrm{SNR}_0\left(\alpha_1^2+\alpha_2^2\right)\left(\beta + (1-\beta)e^{-\rho}\right). \end{aligned} \tag{28}$$

Thus we have

$$\mathrm{MSE}^{(2)} \times \mathrm{SNR}^{(2)} = \left(\beta\rho e^{-\rho} + e^{-\rho} + \frac{1}{\beta}(1-(1+\rho)e^{-\rho})\right)\left(\beta + (1-\beta)e^{-\rho}\right). \tag{29}$$

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 $\alpha_1^* \approx 0.7319$, $\alpha_2^* \approx 0.9550$, and substituting for the optimal values of $\alpha_1^*, \alpha_2^*, \beta^*, \rho^*$ in (28) and multiplying by SNR_0, yields

$$\mathrm{MSE}^{(2)}_{\min} \times \mathrm{SNR}_0 \approx 0.7143. \tag{30}$$

This translates to a 28.47% improvement in MSE performance or a 1.5dB savings in terms of SNR. Figure 3 plots the reduction in MSE for varying values of ρ at optimal α_1^* and the reduction in MSE for varying values of α_1 at optimal ρ^* using simulation (dotted) and theoretically (solid) is shown in Fig. 4. The theoretical curves in both the figures are obtained by evaluating the expression for $\mathrm{MSE}^{(2)}$ in (28) for various values of α_1 and ρ, where α_2 is chosen to satisfy the SNR constraint with equality i.e., $\mathrm{SNR}^{(2)} = \mathrm{SNR}_0$. For the simulation curves, we generate 10000 samples of $\mathbf{y}_1$ in (13) using samples of $\mathbf{x}_1$ obtained from (24) and $\mathbf{n}_1$, where 10000 samples of $\mathbf{n}_1$ and $\mathbf{n}_2$ are generated from $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$. Using samples of $\mathbf{y}_1$, we generate 10000 samples of $\mathbf{x}_2$ from (25) and then obtain 10000 samples of $\mathbf{y}_2$ in (14) using samples of $\mathbf{x}_2$ and $\mathbf{n}_2$. We then obtain an estimate of $\mathrm{MSE}^{(2)}$ by computing the expected value in (16) through numerical averaging over the 10000 realizations. The suboptimal solution to the energy design is shown in Fig. 2 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 two sections, 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 (16) under this strict energy constraint is presented in the following section. This problem was originally addressed in the context of radar imaging in our original paper [42].

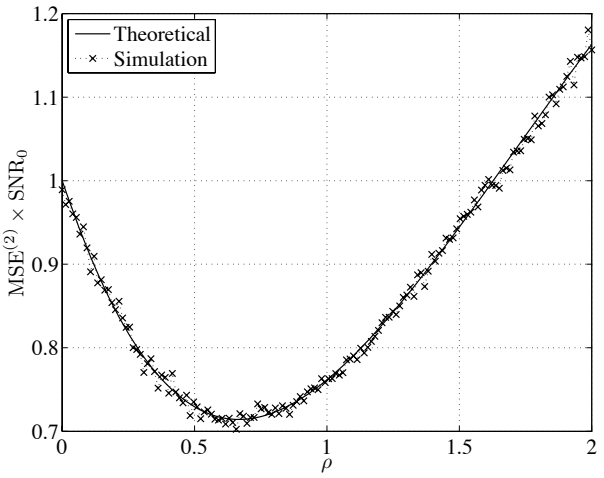


Figure 3: Theoretical versus simulation results for suboptimal strategy. Reduction in MSE for varying values of ρ at optimal $\alpha_1^* = 0.7319$.

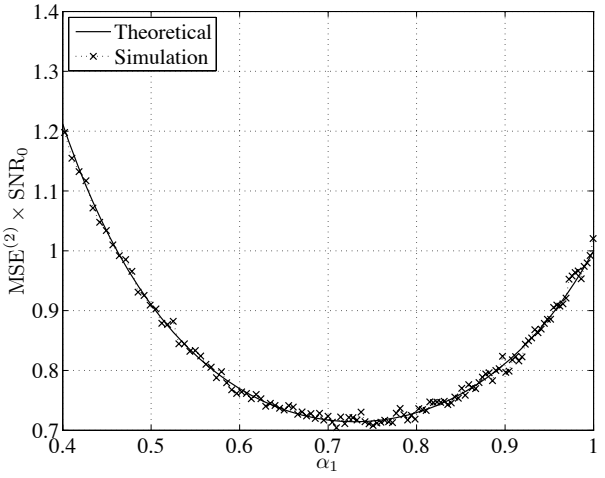


Figure 4: Theoretical versus simulation results for suboptimal strategy. Reduction in MSE for varying values of α_1 at optimal $\rho^* = 0.675$.

5 Strict Energy Constraint Solution

The strict energy constraint for a two-step procedure is $\|\mathbf{x}_1\|^2 + \|\mathbf{x}_2\|^2 \leq E_0$. The MSE for the two-step process given by (16) can be rewritten as

$$\begin{aligned} \mathrm{MSE}^{(2)} &= \sigma^2 \left\{ \frac{1}{(\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2)} - \frac{\|\mathbf{h}_1(\mathbf{x}_1)\|^2(1-|\tilde{n}_1|^2)}{(\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2)^2} \right\} \\ &= \sigma^2(\frac{1}{f} - \frac{c}{f^2}), \end{aligned} \tag{31}$$

where $f = (\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2)$, $c = \|\mathbf{h}_1(\mathbf{x}_1)\|^2(1-|\tilde{n}_1|^2)$, and $\tilde{n}_1$ defined below (20) is complex Gaussian with zero mean and unit variance. Let $\|\mathbf{x}_1\|^2 = E_1$ and $\|\mathbf{x}_2\|^2 = E_2$ such that $E_1 + E_2 \leq E_0$.

Minimizing the MSE with respect to $\mathbf{x}_2(\mathbf{y}_1)$ is equivalent to minimizing with respect to f. f is a function of $\mathbf{x}_1$ and $\mathbf{x}_2(\mathbf{y}_1)$. Since the squared norm of $\mathbf{x}_1$ and $\mathbf{x}_2$ is limited to E_1 and E_2 respectively, the support of f is restricted to $[f_{\min}, f_{\max}]$ obtained by minimizing and maximizing with respect to $\mathbf{x}_2$ respectively:

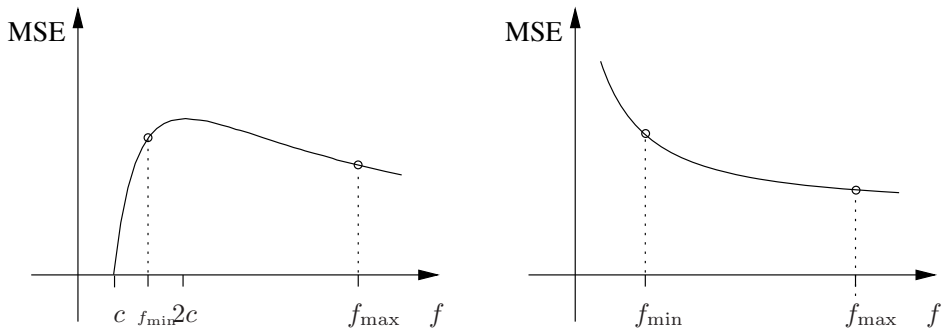
$$\begin{aligned} f_{\min} &= E_1^{'}\lambda_m \quad \text{at} \quad \mathbf{x}_2 = \mathbf{v}_{m\perp}\sqrt{E_2} \\ f_{\max} &= E^{'}\lambda_m \quad \text{at} \quad \mathbf{x}_2 = \mathbf{v}_m\sqrt{E_2}, \end{aligned}$$

where $E_1^{'} = \frac{\|\mathbf{h}_1(\mathbf{x}_1)\|^2}{\lambda_m}$, $E^{'} = E_1^{'} + E_2$, and $\mathbf{v}_{m\perp}$ is an unit norm vector in the perpendicular space of $\boldsymbol{H}_1^H\boldsymbol{H}_1$, i.e., $\mathbf{h}_1(\mathbf{v}_{m\perp}) = 0$. When $\boldsymbol{H}_1^H\boldsymbol{H}_1$ is full rank, then there exists no vector in the orthogonal space, in which case an alternate solution, $E_2 = 0$ can be used to achieve $f_{\min}$. Since $E_2 = 0$ satisfies the energy constraint with inequality and consumes minimal energy, we take $E_2 = 0$ as the optimal solution to achieving $f_{\min}$.

Note that from Fig. 5, the $\mathrm{MSE}^{(2)}(f)$ is either, monotonically increasing from $f = 0$ to $2c$ and decreasing from $f = 2c$ to ∞ for $c > 0$, or strictly decreasing for $c \leq 0$. Since no local minimum exists, the minimum MSE will always occur at the end points of the support of f. Therefore, to minimize the MSE, we simply need to compare the $\mathrm{MSE}^{(2)}$ values at $f_{\min}$ and $f_{\max}$. The optimal $\mathbf{x}_2$ is

$$\begin{aligned} \mathbf{x}_2(\mathbf{y}_1) &= \arg\max_{\mathbf{x}_2}(f)\ \mathbf{I}\left(\mathrm{MSE}^{(2)}(f_{\max}) \leq \mathrm{MSE}^{(2)}(f_{\min})\right) \\ &\quad + \arg\min_{\mathbf{x}_2}(f)\ \mathbf{I}\left(\mathrm{MSE}^{(2)}(f_{\min}) < \mathrm{MSE}^{(2)}(f_{\max})\right) \end{aligned}$$

Since $\mathbf{x}_2 = \mathbf{v}_m\sqrt{E_2}$ maximizes f, $\mathbf{x}_2 = \mathbf{v}_{m\perp}\sqrt{E_2}$ or $E_2 = 0$ minimizes f, and $\mathrm{MSE}^{(2)}(f_{\max}) \leq \mathrm{MSE}^{(2)}(f_{\min})$ is equivalent to $|\tilde{n}_1|^2 \geq \rho$, we conclude



(a) MSE(f) for $c > 0$ (b) MSE(f) for $c \leq 0$.

Figure 5: Typical plots of the MSE as a function of f.

that the optimal $\mathbf{x}_2(\mathbf{y}_1)$ is

$$\mathbf{x}_2(\mathbf{y}_1) = \mathbf{v}_m \sqrt{E_2}\mathbf{I}\left(|\tilde{n}_1|^2 \geq \rho\right) + \mathbf{v}_{m\perp}\sqrt{E_2}\mathbf{I}\left(|\tilde{n}_1|^2 < \rho\right), \tag{32}$$

or equivalently,

$$\mathbf{x}_2(\mathbf{y}_1) = \mathbf{v}_m \sqrt{E_2}\mathbf{I}\left(|\tilde{n}_1|^2 \geq \rho\right), \tag{33}$$

where $\rho = \frac{E_1^{'}}{2E_1^{'}+E_2}$ and $\mathbf{I}(\cdot)$ is the indicator function. This solution implies that when the actual realization of the noise along $\mathbf{h}_1$ in the first transmission is small enough there is no advantage in using the measurement from the second step. Therefore, we transmit $\mathbf{x}_2 \propto \mathbf{v}_{m\perp}$, which makes the overall estimator only a function of the first measurement, or not transmit at the second step by having $E_2 = 0$. When the actual realization of the noise along $\mathbf{h}_1$ in the first transmission is not small enough, there is some merit in incorporating the information from the second measurement and therefore we select $\mathbf{x}_2 \propto \mathbf{v}_m$.

Substituting for $\mathbf{x}_2(\mathbf{y}_1)$ from (33) into (31), we obtain

$$\mathrm{MSE}^{(2)} = \frac{\sigma^2}{E^{'}\lambda_{\mathrm{m}}}\left\{\mathbf{I}\left(|\tilde{n}_1|^2 \geq \rho\right)\left(\frac{E_2}{E^{'}} + \frac{E_1^{'}|\tilde{n}_1|^2}{E^{'}}\right) + \mathbf{I}\left(|\tilde{n}_1|^2 < \rho\right)\left(\frac{E^{'}|\tilde{n}_1|^2}{E_1^{'}}\right)\right\}$$

where $\tilde{n}_1 \sim \mathcal{CN}(0,1)$. Taking the expectation over $\tilde{n}_1$, the $\mathrm{MSE}^{(2)}$ simplifies to

$$\mathrm{MSE}^{(2)} = \frac{\sigma^2}{E^{'}\lambda_{\mathrm{m}}}\left\{e^{-\rho}\frac{1-2\rho}{1-\rho} + \frac{\rho(1+\rho)}{1-\rho}e^{-\rho} + \frac{1-\rho}{\rho}(1-e^{-\rho}(1+\rho))\right\}. \tag{34}$$

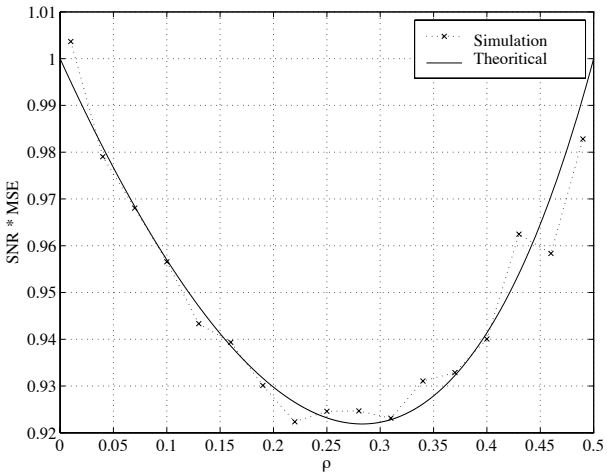


Figure 6: $\text{MSE}^{(2)} \times \text{SNR}^{(2)}$ vs. ρ.

We know that $E^{'} = \frac{\|\mathbf{h}_1(\mathbf{x}_1)\|^2}{\lambda_m} + E_2$. $\text{MSE}^{(2)}$ is minimized when $E^{'}$ is maximized which happens when $\mathbf{x}_1 = \mathbf{v}_m\sqrt{E_1}$. The value of ρ that minimizes the $\text{MSE}^{(2)}$ is given by $\rho^* \approx 0.2831$. This implies that the optimal amount of energy allocated at the first stage is $E_1^* \approx 0.395E_0$ and the remaining energy, $E_2^* \approx 0.605E_0$ is used at the second stage. The minimum MSE will be given by,

$$\text{MSE}^{(2)} \approx \frac{1}{\text{SNR}^{(2)}}(0.9283). \tag{35}$$

We plot the numerical $\text{MSE}^{(2)}$ and the bias as a function of ρ and the exact $\text{MSE}(\rho)$ in Fig. 6 and 7 respectively. It is in fact easy to show that the bias of optimal estimator is zero. The simulation curve was generated by using the design of $\mathbf{x}_2$ given in by generating random 100000 samples of $\mathbf{y}_1$ and $\mathbf{y}_2$. The $\text{MSE}^{(2)}$ was then evaluated by numerically evaluating the expected value in (16). The theoretical curve was generated by evaluating the expression for $\text{MSE}^{(2)}$ in (34). We observe that the simulation results for the $\text{MSE}^{(2)}$ and the bias agree with their analytical equivalents.

The optimal solution satisfies the the strict energy constraint with inequality but the average energy used is only $E_0(\alpha_1^{*2} + \alpha_2^{*2}e^{-\rho^*}) \approx 0.8550E_0$. 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 Section 4. In the following section, we design a θ_1-independent design strategy that achieves the opti-

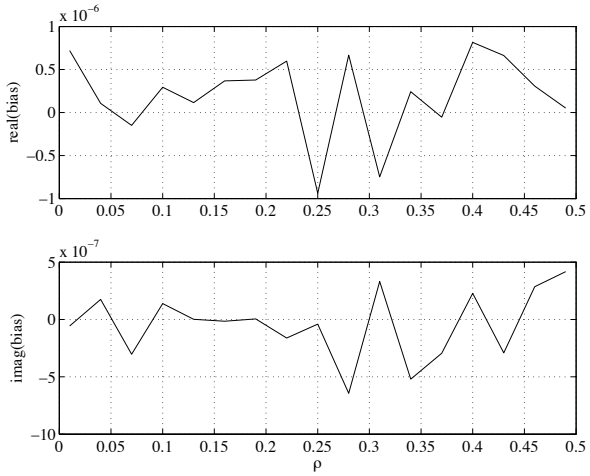


Figure 7: Re(Bias) vs. ρ.

mal performance asymptotically and allows for any peak power constraint in the design.

6 Parameter Independent Two-step Design Strategy

6.1 Problem Statement

Consider the optimal design for the two-step procedure

$$\begin{aligned} \mathbf{x}_1 &= \sqrt{E_0}\ \alpha_1^* \mathbf{v}_{\mathrm{m}} \\ \mathbf{x}_2 &= \sqrt{E_0}\ \alpha_2^*(\tilde{n}_1(\mathbf{y}_1;\theta_1))\mathbf{v}_{\mathrm{m}} = \sqrt{E_0}\ \alpha_2^* \left(\left| \frac{\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\|} \frac{(\mathbf{y}_1 - \sqrt{E_0}\alpha_1 \mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\theta_1)}{\sigma} \right| \right) \mathbf{v}_{\mathrm{m}}. \end{aligned} \tag{36}$$

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 (36) 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., $\theta_1 \in [\theta_{\min}, \theta_{\max}]$, $\theta_{\min}, \theta_{\max} \in \mathbb{R}$.

6.2 Solution

We describe the intuition behind the proposed solution in this subsection. The details of the proof are given in Appendix 12. 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 (36). The resulting suboptimal design is

$$\mathbf{x}_1 = \sqrt{E_0}\ \alpha_1^* \mathbf{v}_{\mathrm{m}} \tag{37}$$

$$\mathbf{x}_2 = \sqrt{E_0}\ \alpha_2^* \left(\left| \frac{\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\|} \frac{(\mathbf{y}_1 - \sqrt{E_0}\alpha_1^* \mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\theta_g)}{\sigma} \right| \right) \mathbf{v}_{\mathrm{m}} = \sqrt{E_0}\ \alpha_2^* \left(|\tilde{n}_1 + z|\right)\ \mathbf{v}_{\mathrm{m}}, \tag{38}$$

where

$$z = \frac{\alpha_1^* \sqrt{E_0} \|\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\|}{\sigma} (\theta_1 - \theta_g) = \alpha_1^* \sqrt{\mathrm{SNR}_0} (\theta_1 - \theta_g) \tag{39}$$

and $\tilde{n}_1$, which is defined below (20) is $\mathcal{CN}(0,1)$. Substituting the above suboptimal solution in the expression for $\mathrm{MSE}^{(N)} \times \mathrm{SNR}^{(N)}$ in (11) and simplifying, we obtain

$$\mathrm{MSE}^{(2)}(z) \times \mathrm{SNR}^{(2)}(z) = \eta(z) = \mathrm{E}\left[\frac{\alpha_1^{*2} |\tilde{n}_1|^2 + \alpha_2^{*2}(|\tilde{n}_1 + z|)}{(\alpha_1^{*2} + \alpha_2^{*2}(|\tilde{n}_1 + z|))^2} \right] \mathrm{E}\left[\alpha_1^{*2} + \alpha_2^{*2}(|\tilde{n}_1 + z|) \right]. \tag{40}$$

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 $\eta(0) = \eta^* = \mathrm{MSE}^{(2)}_{\mathrm{min}} \times \mathrm{SNR}_0 \approx 0.68$, 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. 9 and is shown through an illustration in Fig. 10. Any peak power constraint can also be satisfied using the $N \times 2$-step strategy by choosing a sufficiently large N.

Figure 8 shows $\eta(z)$ in (40) 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

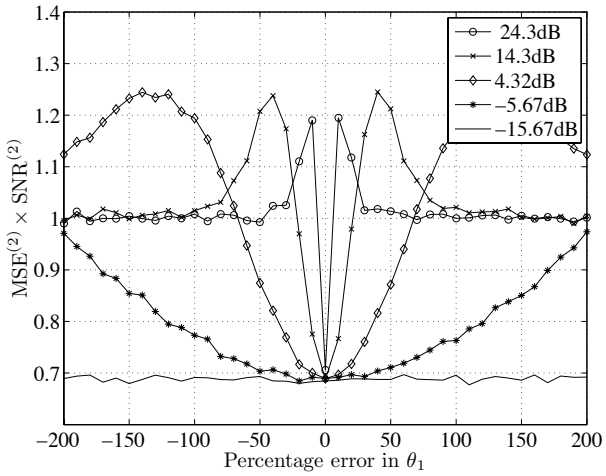


Figure 8: Plot of reduction in MSE versus percentage error in guess of parameter of θ_1 for various SNR.

two-step strategy defined by equations (37) and (38) performs worse than a single step strategy with signal-to-noise ratio SNR_0. This is because the solution presented in (37) and (38) 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 (39) 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 2.1. Thus we observe that the performance of the two-step strategy tends to 1 for large $|\theta_1 - \theta_g|$. The most important information in the plot, however, is the performance of the two-step strategy under low SNR 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.

- *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* (37) *and* (38) *respectively, i.e., use energy E_0/N in each of the N experiments.*
  - The SNR of the $2N$-step procedure is $\mathrm{SNR}^{(2N)}(z) = N\mathrm{SNR}^{(2),1}(z) = \mathrm{SNR}^{(2)}(z/\sqrt{N})$ where $\mathrm{SNR}^{(2),k}$ is the SNR of the k^{th} two-step experiment. The first equality follows from the fact that $\{\mathrm{SNR}^{(2),k}\}_{k=1}^{N}$ 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 $\mathrm{MSE}^{(2N)}(z) = \frac{1}{N}\mathrm{MSE}^{(2),1}(z) = \mathrm{MSE}^{(2)}(z/\sqrt{N})$, where $\mathrm{MSE}^{(2),k}$ is the MSE of each two-step estimator $\hat{\theta}_1^{(2),k}$. The first equality follows from the fact that $\{\mathrm{MSE}^{(2),k}\}_{k=1}^{N}$ 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 $\mathrm{MSE}^{(2N)}(z) \times \mathrm{SNR}^{(2N)}(z) = \mathrm{MSE}^{(2)}(z/\sqrt{N})\mathrm{SNR}^{(2)}(z/\sqrt{N})$. As $N \to \infty$, $z/\sqrt{N} \to 0$ and $\mathrm{MSE}^{(2N)}(z) \times \mathrm{SNR}^{(2N)}(z) \to \eta^*$, i.e., minimal MSE is achieved. The details of the proof can be found in Appendix 12.

Figure 9: Description of the $N\times$ two-step procedure.

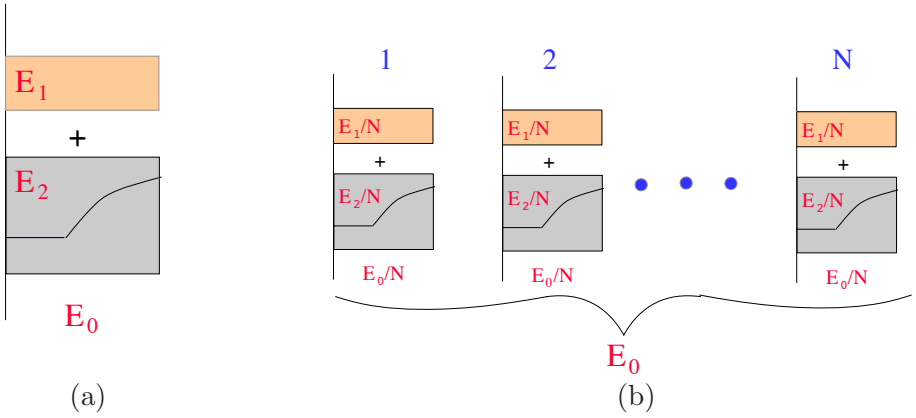


Figure 10: Illustration of the $N\times$ two-step procedure: the omniscient optimal two-step procedure, where energy E_1 is allocated to the first step and E_2 is chosen optimally at the second step based on the past measurements, is shown in Fig. (a). Figure (b) illustrates the $N \times 2$-step procedure, where N independent two-step experiments are performed with the energy design as the optimal two-step energy allocation strategy scaled through $1/N$ but with θ_g replacing θ_1. By averaging the estimates of the N two-step estimators, we asymptotically achieve optimal performance as $N \to \infty$.

7 Design of N-step procedure

In Sections 3 and 6, 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 section, 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 (3). 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

$$\begin{aligned} \alpha_1 &= A_1 && (41) \\ \alpha_i &= A_i I\left(\frac{|\sum_{j=1}^{i-1} \mathbf{h}_1(\mathbf{x}_j)^H \mathbf{n}_j|^2}{\sum_{j=1}^{i-1} \|\mathbf{h}_1(\mathbf{x}_j)\|^2 \sigma^2} \geq \rho_i\right), \quad i \geq 2, && (42) \end{aligned}$$

where $\{A_i, \rho_i\}$ are design parameters. This approximate solution is motivated from the suboptimal thresholding solution to the two-step case derived in Section 4. 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,

$$\alpha_2 = A_2 I\left(\left|\frac{\mathbf{h}_1(\mathbf{x}_1)^H}{\|\mathbf{h}_1(\mathbf{x}_1)\|}\frac{\mathbf{n}_1}{\sigma}\right|^2 \geq \rho_2\right) = A_2 I\left(\left|\frac{\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\|}\frac{\mathbf{n}_1}{\sigma}\right|^2 \geq \rho_2\right) = A_2 I\left(|\tilde{n}_1|^2 \geq \rho_2\right), \tag{43}$$

where $\tilde{n}_i = \frac{\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\|}\frac{\mathbf{n}_i}{\sigma}$ are i.i.d complex Gaussian random variables with zero mean and unit variance. The amplitude at the third stage simplifies to

$$\begin{aligned} \alpha_3 &= A_3\left(\frac{|\mathbf{h}_1(\mathbf{x}_1)^H\mathbf{n}_1 + \mathbf{h}_1(\mathbf{x}_2)^H\mathbf{n}_2|^2}{\sigma^2\left(\|\mathbf{h}_1(\mathbf{x}_1)\|^2 + \|\mathbf{h}_1(\mathbf{x}_2)\|^2\right)} \geq \rho_3\right) \\ &= A_3 I\left(\frac{|A_1\tilde{n}_1 + A_2\tilde{n}_2|^2}{|A_1|^2 + |A_2|^2} \geq \rho_3\right) I\left(|\tilde{n}_1|^2 \geq \rho_2\right) + A_3 I\left(|\tilde{n}_1|^2 \geq \rho_2\right) I\left(|\tilde{n}_1|^2 < \rho_2\right) \\ &= A_3 I\left(\frac{|A_1\tilde{n}_1 + A_2\tilde{n}_2|^2}{|A_1|^2 + |A_2|^2} \geq \rho_3\right) I\left(|\tilde{n}_1|^2 \geq \rho_2\right). \end{aligned} \tag{44}$$

Following the same procedure, we simplify α_4 as

$$\alpha_4 = A_4 I\left(\frac{|A_1\tilde{n}_1 + A_2\tilde{n}_2 + A_3\tilde{n}_3|^2}{|A_1|^2 + |A_2|^2 + |A_3|^2} \geq \rho_4\right) I\left(\frac{|A_1\tilde{n}_1 + A_2\tilde{n}_2|^2}{|A_1|^2 + |A_2|^2} \geq \rho_3\right) I\left(|\tilde{n}_1|^2 \geq \rho_2\right). \tag{45}$$

Thus, a general expression for α_i can be written as

$$\alpha_i = A_i \prod_{s=1}^{i-1} I\left(|w_s|^2 \geq \rho_{s+1}\right), \tag{46}$$

where w_s is defined in (134). 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 $\mathcal{G}_N = \text{MSE}^{(N)} \times \text{SNR}^{(N)}$ which from Appendix 13 is given by

$$\text{MSE}^{(N)} \times \text{SNR}^{(N)}(\mathbf{A}, \boldsymbol{\rho}) = \left(\sum_{i=1}^{N-1} \frac{T_i}{Q_i} + \frac{\tilde{T}_N}{Q_N}\right)\left(\sum_{i=1}^{N-1} Q_i P_i + Q_N \tilde{P}_N\right), \tag{47}$$

where $\mathbf{A} = [A_1, \ldots, A_N], \boldsymbol{\rho} = [\rho_1, \ldots, \rho_N]$, Q_i, T_i, P_i are defined in (136), (137) and (139) respectively.

There is no closed-form solution to this $2N$ dimensional optimization. Instead 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 4 and d is chosen to satisfy the average energy constraint. 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. 12. The theoretical performance curve was generated by evaluating the expressions for T_i and P_i in (137) and (139) using numerical integration and substituting them in (47) while the simulations are generated by sampling the distributions of $\{\mathbf{y}_i\}_{i=1}^N$ by first generating samples of $\{\alpha_i\}_{i=1}^N$ followed by an empirical estimate of the MSE. By designing this N-step procedure, we are essentially altering the Gaussian statistics of the measurement noise to obtain improvements in performance. In Fig. 11, we illustrate how the distribution of the estimation residuals changes with the number of the steps. We would like to point out that the simulation curve appears smoother than the theory curve in Fig. 12 as evaluating (137) and (139) involved high order integration. We see that in 50 steps, we are able to achieve gains of more than 5dB. In Section 6, 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

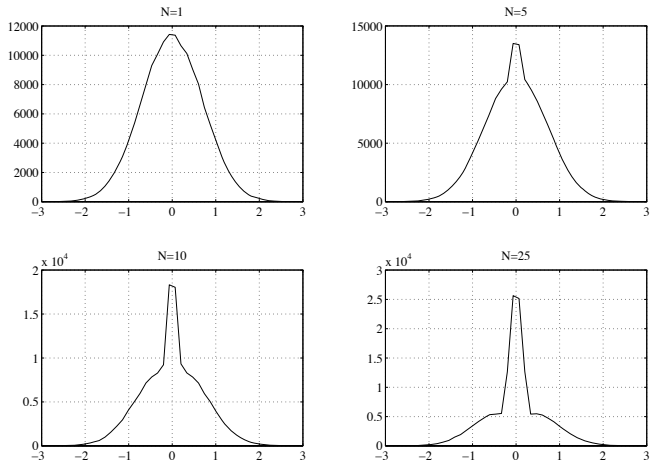


Figure 11: Distribution of noise versus number of steps.

error of guessing θ_1 has a reduced effect on the overall performance. We demonstrate the achievability of performance for any N-step design in the following subsection.

7.0.1 Achievability of performance of any omniscient N-step design

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 (5).

Theorem 7.1. *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* - $\mathrm{E}\left[\sum_{i=1}^N \|\mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1}; \theta_1)\|^2\right] \leq E_0$.
- *Continuity - The design vector $\mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1}; \theta_1)$ is a continuous function of $\{\mathbf{y}_j\}_{j=1}^{i-1}$ or can assume the form of a thresholding function in (42).*

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

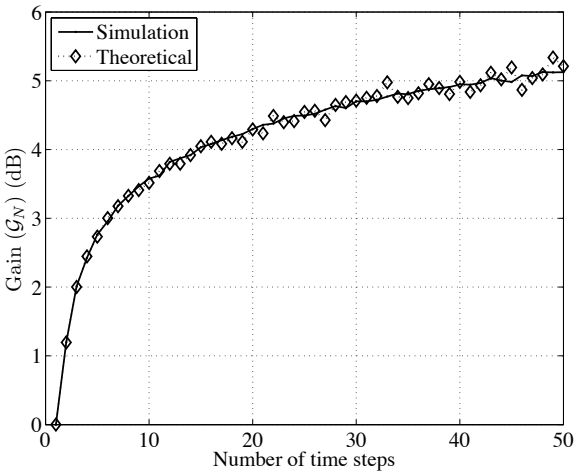


Figure 12: Plot of gains obtained through suboptimal N-step procedure as a function of N through theory and simulations.

Proof. The proof is similar to the $N \times 2$-step strategy presented in Section 6, where the actual value of θ_1 in the optimal solution is replaced with a guess of θ_1. Refer to Appendix 16 for details. □

8 Sequential Design for Vector Parameters

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

$$\mathbf{y}_i \;=\; \mathbf{H}(\mathbf{x}_i(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1}))\boldsymbol{\theta} + \mathbf{n}_i, \quad i = 1, 2, \ldots, N, \tag{48}$$

where $\boldsymbol{\theta}$ is an M-element vector, $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{R_n})$, 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.

8.1 Worst Case Error - Min Max Approach

The component wise MSE for estimating specific parameters is given by the diagonal elements of the matrix 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 $\mathbf{u}_{\rm m}$ is any unit norm vector independent of past measurements, e.g., $\mathbf{u}_{\rm m}$ is chosen to minimize the one-step MSE. The ML estimate for a one-step process with energy E_0 is given by

$$\widehat{\boldsymbol{\theta}}^{(1)} = \frac{1}{\sqrt{E_0}} \mathbf{W}_{\mathbf{u}_{\rm m}} \mathbf{H}(\mathbf{u}_{\rm m})^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{y}_1 \tag{49}$$

and its corresponding MSE is

$$\text{MSE}^{(1)} = \frac{1}{E_0} \mathbf{W}_{\mathbf{u}_{\rm m}}, \tag{50}$$

where $\mathbf{W}_{\mathbf{u}_{\rm m}} = (\mathbf{H}(\mathbf{u}_{\rm m})^H \mathbf{R}_{\mathbf{n}}^{-1} \mathbf{H}(\mathbf{u}_{\rm m}))^{-1}$. Define $\Phi(\mathbf{u}, \text{MSE}) = \mathbf{u}^H \text{MSE} \mathbf{u}$.

$$WC\text{-MSE} = \max_i \mathbf{e}_i^H \text{MSE} \mathbf{e}_i = \max_i \Phi(\mathbf{e}_i, \text{MSE}), \tag{51}$$

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

$$WC\text{-MSE}^{(1)} = \max_i \Phi(\mathbf{e}_i, \text{MSE}^{(1)}) = \Phi(\mathbf{e}_{i^*}, \text{MSE}^{(1)}), \tag{52}$$

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

$$\Phi(\mathbf{u}, \text{MSE}^{(1)}) = \frac{1}{E_0} \mathbf{u}^H \mathbf{W}_{\mathbf{u}_{\rm m}} \mathbf{u}. \tag{53}$$

The set of observations for the two-step process are

$$\mathbf{y}_1 = \sqrt{E_0} \alpha_1 \mathbf{H}(\mathbf{u}_{\rm m}) \boldsymbol{\theta} + \mathbf{n}_1 \tag{54}$$

$$\mathbf{y}_2 = \sqrt{E_0} \alpha_2(\mathbf{y}_1) \mathbf{H}(\mathbf{u}_{\rm m}) \boldsymbol{\theta} + \mathbf{n}_2. \tag{55}$$

For a two-step procedure, we need to design α_1 and $\alpha_2(\mathbf{y}_1)$ to minimize $WC\text{-MSE}^{(2)}$. From (141) in Appendix 14, we have

$$\Phi(\mathbf{u}, \text{MSE}^{(2)}) = \Phi(\mathbf{u}, \text{MSE}^{(1)}) \, \text{E} \left[\frac{\alpha_1^2 \left| \tilde{n}_1(\mathbf{y}_1; \boldsymbol{\theta}) \right|^2 + \alpha_2^2(\tilde{n}_1(\mathbf{y}_1; \boldsymbol{\theta}))}{(\alpha_1^2 + \alpha_2^2(\tilde{n}_1(\mathbf{y}_1; \boldsymbol{\theta}))^2} \right], \tag{56}$$

where $\tilde{n}_1(\mathbf{y}_1; \boldsymbol{\theta}) = \frac{\mathbf{u}^H \mathbf{W}_{\mathbf{u}_{\rm m}} \mathbf{H}(\mathbf{u}_{\rm m})^H \mathbf{R}_{\mathbf{n}}^{-1} (\mathbf{y}_1 - \mathbf{H}(\mathbf{x}_1)\boldsymbol{\theta})}{\sqrt{\mathbf{u}^H \mathbf{W}_{\mathbf{u}_{\rm m}} \mathbf{u}}}$ is a complex normal random variable with zero mean and unit variance. The error in (56) 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 3. 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 $\Phi(\mathbf{u}, \mathrm{MSE}^{(2)}) \approx 0.6821\ \Phi(\mathbf{u}, \mathrm{MSE}^{(1)})$. It follows that

$$WC\text{-}\mathrm{MSE}^{(2)} = \Phi(\mathbf{e}_{i^*}, \mathrm{MSE}^{(2)}) \approx 0.6821\ \Phi(\mathbf{e}_{i^*}, \mathrm{MSE}^{(1)}) = 0.6821\ WC\text{-}\mathrm{MSE}^{(1)} \tag{57}$$

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 6 [43]. The reduction in MSE in (57) holds for any M, the number of unknown parameters, as i^*, the index of the worst case error, can always be computed from (52) and (53) for any $M \in \mathbb{N}$. A similar result can be derived for the N-step procedure.

8.2 Trace Criteria

For the multiple parameter case, the MSE is a matrix and we consider the trace as a measure of performance i.e., $\min_{\{\mathbf{x}_i\}_{i=1}^N} \mathrm{tr}(\mathrm{MSE}^{(N)}(\boldsymbol{\theta}))$, where $\mathrm{tr}(\cdot)$ denotes the trace. So far, we considered the problem of optimal and suboptimal strategies for energy allocation in an N-time step procedure. We assumed in our analysis that the waveform transmitted is the one-step optimal derived in Section 2.1. For the purposes of the trace criterion in the vector parameter case, we consider two possible strategies: first, performing N-step energy allocation under the constraint that the waveform transmit at every time instant is the one-step optimal for estimating the vector parameters. The second strategy is to provide a waveform and energy allocation simultaneously. We present the energy allocation procedure in the following section.

8.2.1 Energy design under fixed waveforms

The trace of the MSE matrix can be written in the following form,

$$\mathrm{tr}(\mathrm{MSE}) \quad = \quad \sum_{i=1}^{M} \mathbf{e}_i^H \mathrm{MSE} \mathbf{e}_i = \sum_{i=1}^{M} \Phi(\mathbf{e}_i, \mathrm{MSE}), \tag{58}$$

In this analysis, we will assume that all the elements of the matrix $\mathbf{H}$ defined in (48) can be controlled by the design vector $\mathbf{x}_i$ at every time instant. In other words, we assume the following statistical model

$$\mathbf{y}_i = \alpha_i \sqrt{E_0} \widetilde{\mathbf{X}}_i \boldsymbol{\theta} + \mathbf{n}_i, i = 1, 2, \ldots, N, \tag{59}$$

where $\widetilde{\mathbf{X}}_i$ is the $K \times M$ input design matrix and $\mathbf{n}_i \sim \mathcal{CN}(\mathbf{0}, \mathbf{R_n})$. The one-step estimator for $\boldsymbol{\theta}$ in this model is given by

$$\widehat{\boldsymbol{\theta}}^{(1)} = \frac{1}{\sqrt{E_0}}(\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1} \widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \mathbf{y}_1 \tag{60}$$

and the corresponding MSE can be derived using a similar derivation to (50) as

$$\text{MSE}^{(1)} = \frac{1}{E_0}(\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1}. \tag{61}$$

The trace of the $\text{MSE}^{(1)}$ matrix can then be written as

$$\text{tr}(\text{MSE}^{(1)}) = \frac{1}{E_0}\text{tr}\left\{(\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1}\right\} = \frac{1}{E_0}\sum_{i=1}^{M} \mathbf{e}_i^H (\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1} \mathbf{e}_i \tag{62}$$

For a two-step process, the set of observations can be written as

$$\mathbf{y}_1 = \alpha_1 \sqrt{E_0} \widetilde{\mathbf{X}}_1 \boldsymbol{\theta} + \mathbf{n}_1 \tag{63}$$

$$\mathbf{y}_2 = \alpha_2(\mathbf{y}_1) \sqrt{E_0} \widetilde{\mathbf{X}}_1 \boldsymbol{\theta} + \mathbf{n}_2, \tag{64}$$

where $\widetilde{\mathbf{X}}_1$ will be chosen as the optimal one-step design and α_1 and $\alpha_2(\mathbf{y}_1)$ are optimal energy allocation design parameters. The two-step ML estimator is given by

$$\widehat{\boldsymbol{\theta}}^{(2)} = \frac{1}{\sqrt{E_0}}(\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1} \widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \left(\frac{\alpha_1 \mathbf{y}_1 + \alpha_2(\mathbf{y}_1)\mathbf{y}_2}{\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)}\right) \tag{65}$$

and the corresponding MSE is

$$\text{MSE}^{(2)} = \frac{1}{E_0}\text{E}_{\mathbf{n}_1}\left[(\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1} \widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \left(\frac{\alpha_1^2 \mathbf{n}_1 \mathbf{n}_1^H + \alpha_2^2(\mathbf{y}_1)\mathbf{R_n}}{\left(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)\right)^2}\right) \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1 (\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1}\right]. \tag{66}$$

The trace of the $\text{MSE}^{(2)}$ can be evaluated as

$$\begin{aligned}\text{tr}(\text{MSE}^{(2)}) &= \frac{1}{E_0}\left(\text{E}\left[\frac{\alpha_1^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2} \mathbf{n}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1 (\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-2} \widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \mathbf{n}_1 \right.\right. \\ &\qquad \left.\left. + \frac{\alpha_2^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2} \text{tr}\left\{(\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1}\right\}\right]\right) \qquad (67) \\ &= \frac{1}{E_0}\left(\text{E}\left[\frac{\alpha_1^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2} \widetilde{\mathbf{n}}_1^H(\mathbf{y}_1; \boldsymbol{\theta}) \mathbf{M} \widetilde{\mathbf{n}}_1(\mathbf{y}_1; \boldsymbol{\theta}) \right.\right. \\ &\qquad \left.\left. + \frac{\alpha_2^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2} \text{tr}\left\{(\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1}\right\}\right]\right), \qquad (68)\end{aligned}$$

where

$$\mathbf{M} = \mathbf{R}_{\mathbf{n}}^{-1/2}\widetilde{\mathbf{X}}_1(\widetilde{\mathbf{X}}_1^H\mathbf{R}_{\mathbf{n}}^{-1}\widetilde{\mathbf{X}}_1)^{-2}\widetilde{\mathbf{X}}_1^H\mathbf{R}_{\mathbf{n}}^{-1/2} \tag{69}$$

and

$$\widetilde{\mathbf{n}}_1(\mathbf{y}_1;\boldsymbol{\theta}) = \mathbf{R}_{\mathbf{n}}^{-1/2}\mathbf{n}_1 = \mathbf{R}_{\mathbf{n}}^{-1/2}(\mathbf{y}_1 - \alpha_1\sqrt{E_0}\widetilde{\mathbf{X}}_1\boldsymbol{\theta}). \tag{70}$$

Since rotation or translation of the Gaussian random vector is still a Gaussian random vector, it follows that $\widetilde{\mathbf{n}}_1$ is i.i.d $\mathcal{CN}(\mathbf{0},\mathbf{I})$. Using circular invariance of trace of a matrix, i.e., $\mathrm{tr}(\mathbf{ABC}) = \mathrm{tr}(\mathbf{CAB})$, we have

$$\mathrm{tr}(\mathbf{M}) \;=\; \mathrm{tr}\left\{\mathbf{R}_{\mathbf{n}}^{-1/2}\widetilde{\mathbf{X}}(\widetilde{\mathbf{X}}^H\mathbf{R}_{\mathbf{n}}^{-1}\widetilde{\mathbf{X}})^{-2}\widetilde{\mathbf{X}}^H\mathbf{R}_{\mathbf{n}}^{-1/2}\right\} = \mathrm{tr}\left\{(\widetilde{\mathbf{X}}^H\mathbf{R}_{\mathbf{n}}^{-1}\widetilde{\mathbf{X}})^{-1}\right\} \tag{71}$$

Define $w_1(\mathbf{y}_1;\boldsymbol{\theta})$ as

$$w_1(\mathbf{y}_1;\boldsymbol{\theta}) = \frac{\widetilde{\mathbf{n}}_1^H(\mathbf{y}_1;\boldsymbol{\theta})\mathbf{M}\widetilde{\mathbf{n}}_1(\mathbf{y}_1;\boldsymbol{\theta})}{\mathrm{tr}(\mathbf{M})}. \tag{72}$$

Substituting (72) and (71) in (68), we obtain

$$\mathrm{tr}(\mathrm{MSE}^{(2)}) = \mathrm{tr}(\mathrm{MSE}^{(1)})\left(\mathrm{E}\left[\frac{\alpha_1^2}{(\alpha_1^2+\alpha_2^2(w_1(\mathbf{y}_1;\boldsymbol{\theta})))^2}w_1(\mathbf{y}_1;\boldsymbol{\theta}) + \frac{\alpha_2^2(w_1)}{(\alpha_1^2+\alpha_2^2(w_1(\mathbf{y}_1;\boldsymbol{\theta})))^2}\right]\right), \tag{73}$$

where $\alpha_2(\mathbf{y}_1)$ is replaced with $\alpha_2(w_1(\mathbf{y}_1;\boldsymbol{\theta}))$ as the dependence of α_2 on $\mathbf{y}_1$ occurs through $w_1(\mathbf{y}_1;\boldsymbol{\theta})$. We will denote $w_1(\mathbf{y}_1;\boldsymbol{\theta})$ and $\widetilde{\mathbf{n}}_1^H(\mathbf{y}_1;\boldsymbol{\theta})$ as w_1 and $\widetilde{\mathbf{n}}_1$ respectively in the remainder of this discussion for convenience.

Let $\mathbf{M} = \mathbf{QDQ}^H$ denote the eigenvalue decomposition of the matrix $\mathbf{M}$, where $\mathbf{D} = [d_1,\ldots,d_N]$ denotes the eigenvalues. Then

$$w_1 \;=\; \frac{\widetilde{\mathbf{n}}_1^H\mathbf{M}\widetilde{\mathbf{n}}_1}{\mathrm{tr}(\mathbf{M})} = \frac{\widehat{\mathbf{n}}_1^H\mathbf{D}\widehat{\mathbf{n}}_1}{\mathrm{tr}(\mathbf{D})} = \frac{\sum_{i=1}^N d_i|\widehat{\mathbf{n}}_{1,i}|^2}{\sum_{i=1}^N d_i} = \sum_{i=1}^N\left(\frac{d_i}{\sum_{i=1}^N d_i}\right)|\widehat{\mathbf{n}}_{1,i}|^2 \tag{74}$$

where $\widehat{\mathbf{n}}_1 = \mathbf{Q}^H\widetilde{\mathbf{n}}_1$ is again $\mathcal{CN}(0,\mathbf{I})$ since $\mathbf{Q}$ is unitary and $\widehat{\mathbf{n}}_{1,i}$ denotes the ith element of the vector $\widehat{\mathbf{n}}_1$. $\{|\widehat{\mathbf{n}}_{1,i}|^2\}_{i=1}^N$ are independent central χ_2^2 random variables, i.e., chi-square random variables with 2 degrees of freedom (exponentially distributed). Hence w_1 is central chi-square mixture with N degrees of freedom with

$$\mathrm{E}\left[w_1\right] \;=\; 1 \tag{75}$$

$$\mathrm{var}(w_1) \;=\; 2\sum_{i=1}^N\left(\frac{d_i}{\sum_{i=1}^N d_i}\right)^2 \tag{76}$$

To find the optimal solution to the energy at the second stage, i.e., $\alpha_2(w_1)$ we need to minimize with respect to $\alpha_2(w_1)$ the function

$$\left(\mathrm{E}\left[\frac{\alpha_1^2}{(\alpha_1^2+\alpha_2^2(w_1))^2} w_1 + \frac{\alpha_2^2(w_1)}{(\alpha_1^2+\alpha_2^2(w_1))^2} \right] \right) \tag{77}$$

subject to the constraint on the average energy $\mathrm{E}\left[\alpha_1^2+\alpha_2^2(w_1)\right] \leq 1$. The Lagrangian, the minimization condition plus the constraint, can be written as

$$\min_{\alpha_1,\alpha_2(w_1)} \mathrm{E}_{w_1}\left[\frac{\alpha_1^2 w_1 + \alpha_2^2(w_1)}{(\alpha_1^2+\alpha_2^2(w_1))^2} \right] + \lambda\left(\alpha_1^2 + \mathrm{E}_{w_1}\left[\alpha_2^2(w_1)\right]\right) \tag{78}$$

$$= \min_{\alpha_1,\alpha_2(w_1)} \frac{1}{\alpha_1^2}\mathrm{E}_{\mathbf{n}_1}\left[\frac{w_1 + \alpha_2^2(w_1)/\alpha_1^2}{(1+\alpha_2^2(w_1)/\alpha_1^2)^2} \right] + \lambda\alpha_1^2\left(1 + \mathrm{E}_{w_1}\left[\alpha_2^2(w_1)/\alpha_1^2\right]\right) \tag{79}$$

$$= \min_{\alpha_1,g(w_1)} \frac{1}{\alpha_1^2}\mathrm{E}_{w_1}\left[\frac{w_1 + g(w_1) - 1}{g^2(w_1)} \right] + \lambda\alpha_1^2\mathrm{E}_{w_1}\left[g(w_1)\right] \tag{80}$$

$$= \min_{\alpha_1,g(w_1)} \frac{1}{\alpha_1^2}\mathrm{E}_{w_1}\left[\frac{1}{g(w_1)} - \frac{1-w_1}{g^2(w_1)} + \lambda^{'} g(w_1) \right], \tag{81}$$

where $g(w_1) = 1 + \alpha_2^2(w_1)/\alpha_1^2$ and $\lambda^{'} = \lambda\alpha_1^2$. Differentiating and setting the derivative with respect to g to zero,

$$g^3 - \frac{1}{\lambda^{'}} g + 2\frac{1-w_1}{\lambda^{'}} \quad = \quad 0 \tag{82}$$

The structure of the optimal design for the multiple parameter case is the same as that of the optimal design for the single parameter case. The only difference between the solutions is that the optimal solution $\alpha_2(w_1)$ is a function of w_1, a χ^2-mixture with N degrees of freedom which reduces to being a exponentially distributed random variable when the number of parameters to be determined is reduced to 1. Furthermore to compute the exact reduction in the MSE, we need to know the general distribution of the χ^2-mixture. Various series expansions have been proposed in the literature for the distribution of a sum of χ^2-random variables: power series [44], χ^2 series [45], improved power series and Laguerre series expansions [46], Laguerre series for non central chi-square sum [47].

Before we proceed to find the optimal sequential design of energy for an N-step process, we solve for the optimal $\widetilde{\mathbf{X}}_1$ for an one-step process which yields the eigenvalues $d_1, d_2, \ldots, d_N$. To find the optimal one-step design vector, we need the following results.

Lemma 8.1. $N \geq M$

Proof. Consider the $M \times M$ matrix $(\widetilde{\mathbf{X}}^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}})$ in the solution to the maximum likelihood estimator of $\boldsymbol{\theta}$ in (49).

$$(\widetilde{\mathbf{X}}^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}) = \left\{ (\widetilde{\mathbf{X}}^H \mathbf{R}_\mathbf{n}^{-1/2})(\widetilde{\mathbf{X}}^H \mathbf{R}_\mathbf{n}^{-1/2})^H \right\}$$

Since for any matrix $\mathbf{A}$, $Rank(\mathbf{A}\mathbf{A}^H) = Rank(\mathbf{A})$, it follows that

$$Rank(\widetilde{\mathbf{X}}^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}) \quad = \quad Rank(\widetilde{\mathbf{X}}^H \mathbf{R}_\mathbf{n}^{-1/2}) = Rank(\widetilde{\mathbf{X}}) = \min(N, M). \quad (83)$$

Since we want the $M \times M$ matrix $(\widetilde{\mathbf{X}}^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}})$ to be invertible, we need $Rank(\widetilde{\mathbf{X}}^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}) = M$. It follows that $N \geq M$ for any M. □

For the rest of the derivation, we consider the following assumptions:

- Without loss of generality we assume $N = M$ and furthermore $\widetilde{\mathbf{X}}$ is full rank.
- The receiver noises are independent and identically distributed.i.e., $\mathbf{R}_\mathbf{n} = \sigma^2 \mathbf{I}$.

Lemma 8.2. *For any positive $m \times m$ definite matrix $\mathbf{A}$, the following inequality holds,*

$$tr(\mathbf{A}^{-1}) \quad \geq \quad \sum_{i=1}^{M} (a_{i,i})^{-1}, \quad (84)$$

where $a_{i,i}$ is the ith diagonal element of A and equality iff $\mathbf{A}$ is diagonal.

Proof. The details of the proof can be found in [48]. □

Theorem 8.3. *$d_i = \sigma^2 N$, $i = 1, 2, \ldots, N$ minimizes $MSE^{(1)}$.*

Proof. The single-stage MSE given in (62) has the form

$$\mathrm{tr}(\mathrm{MSE}^{(1)}) \quad = \quad \frac{1}{E_0} (\widetilde{\mathbf{X}}_1^H \mathbf{R}_\mathbf{n}^{-1} \widetilde{\mathbf{X}}_1)^{-1} = \frac{\sigma^2}{E_0} \mathrm{tr} \left\{ (\widetilde{\mathbf{X}}_1^H \widetilde{\mathbf{X}}_1)^{-1} \right\}, \quad (85)$$

when $\mathbf{R}_\mathbf{n} = \sigma^2 \mathbf{I}$. To obtain the minimum error we need to find $\widetilde{\mathbf{X}}_1$ optimally. Since the energy component was already extracted as a multiplying term of the form $\sqrt{E_0}\alpha_j(\mathbf{y}_1, \ldots, \mathbf{y}_{j-1})$, $\widetilde{\mathbf{X}}_1$ should satisfy $\|\widetilde{\mathbf{X}}_1\|_F = 1$ ($\|\cdot\|_F$ denotes the Frobenius norm). Using Lemma 8.2, it follows that the optimal $\widetilde{\mathbf{X}}_1$ satisfies

$$\widetilde{\mathbf{X}}_1^H \widetilde{\mathbf{X}}_1 = \mathrm{diag}(\|\mathbf{z}_1\|^2, \ldots, \|\mathbf{z}_N\|^2), \quad (86)$$

where $\widetilde{\mathbf{X}}_1 = [\mathbf{z}_1, \ldots, \mathbf{z}_N]$. We use Lagrange multipliers to solve the following optimization problem. The Lagrangian is given by

$$\begin{aligned} L(\widetilde{\mathbf{X}}_1, \lambda) &= \min_{\widetilde{\mathbf{X}}_1} \operatorname{tr}\left\{ \left(\widetilde{\mathbf{X}}_1^H \widetilde{\mathbf{X}}_1 \right)^{-1} \right\} + \lambda(\operatorname{tr}(\widetilde{\mathbf{X}}_1^H \widetilde{\mathbf{X}}_1) - 1) \quad (87) \\ &= \sum_{i=1}^{N} \left(\|\mathbf{z}_i\|^2 \right)^{-1} + \lambda \left(\sum_{i=1}^{N} \|\mathbf{z}_i\|^2 - 1 \right). \quad (88) \end{aligned}$$

Setting $\frac{\partial L(\widetilde{\mathbf{X}}_1, \lambda)}{\partial \mathbf{z}_j} = 0$ for $j = 1, \ldots, N$, we obtain

$$\begin{aligned} \left(\|\mathbf{z}_j\|^2 \right)^{-2} \mathbf{z}_j - \lambda \mathbf{z}_j &= 0 \quad (89) \\ \left(1 - \lambda |\mathbf{z}_j\|^4 \right) \mathbf{z}_j &= 0. \quad (90) \end{aligned}$$

It follows that the solution to $\{\|\mathbf{z}_i\|^2\}_{i=1}^N$ is

$$\|\mathbf{z}_i\|^2 = \sqrt{\frac{1}{\lambda_0}}, \quad i = 1, \ldots, N \quad (91)$$

The optimal λ, $\lambda^* = N^2$ and

$$\|\mathbf{z}_i\|^2 = \frac{1}{N}, \quad i = 1, \ldots, N \quad (92)$$

Hence $\widetilde{\mathbf{X}}_1^H \widetilde{\mathbf{X}}_1 = \frac{1}{N}\mathbf{I}$ and

$$\begin{aligned} \mathbf{M} &= \mathbf{R}_{\mathbf{n}}^{-1/2} \widetilde{\mathbf{X}}_1 (\widetilde{\mathbf{X}}_1^H \mathbf{R}_{\mathbf{n}}^{-1} \widetilde{\mathbf{X}}_1)^{-2} \widetilde{\mathbf{X}}_1^H \mathbf{R}_{\mathbf{n}}^{-1/2} = \sigma^2 \ N^2 \widetilde{\mathbf{X}}_1 \widetilde{\mathbf{X}}_1^H \\ \left(\mathbf{M} - \sigma^2 N \mathbf{I} \right) \widetilde{\mathbf{X}}_1 &= 0 \\ \mathbf{M} &= \sigma^2 N \mathbf{I} \quad (93) \end{aligned}$$

It follows that all eigenvalues $d_1, \ldots, d_N$ of the matrix $\mathbf{M}$ are equal to $\sigma^2 \ N$. □

Since all eigenvalues are equal, the expression for w_1 in (74) can be simplified to

$$w_1 = \sum_{i=1}^{N} \left(\frac{d_i}{\sum_{i=1}^{N} d_i} \right) |\widehat{\mathbf{n}}_{1,i}|^2 = \frac{1}{N} \sum_{i=1}^{N} |\widehat{\mathbf{n}}_{1,i}|^2. \quad (94)$$

From Appendix 15, we see that w_1 is a Gamma distribution whose probability density function and cumulative distribution function as

$$f_{w_1}(y) = \frac{N^N}{(N-1)!} y^{n-1} e^{-Ny}, \ y \geq 0 \quad (95)$$

$$F_{w_1}(y) = 1 - e^{-Ny} \sum_{j=0}^{N} \frac{(Ny)^j}{j!}, \quad (96)$$

M	α_1^*	λ'^*	M	α_1^*	λ'^*	M	α_1^*	λ'^*
1.0000	0.7427	0.2077	11.0000	0.8781	0.5652	21.0000	0.9002	0.6373
2.0000	0.7910	0.3163	12.0000	0.8817	0.5766	22.0000	0.9003	0.6373
3.0000	0.8152	0.3787	13.0000	0.8817	0.5766	23.0000	0.9041	0.6502
4.0000	0.8295	0.4186	14.0000	0.8853	0.5883	24.0000	0.9041	0.6502
5.0000	0.8417	0.4535	15.0000	0.8890	0.6002	25.0000	0.9042	0.6502
6.0000	0.8512	0.4816	16.0000	0.8926	0.6123	26.0000	0.9080	0.6634
7.0000	0.8577	0.5012	17.0000	0.8927	0.6123	27.0000	0.9080	0.6634
8.0000	0.8643	0.5217	18.0000	0.8964	0.6247	28.0000	0.9080	0.6634
9.0000	0.8677	0.5323	19.0000	0.8965	0.6247	29.0000	0.9119	0.6768
10.0000	0.8746	0.5540	20.0000	0.9002	0.6373	30.0000	0.9119	0.6768

Table 2: Optimal values of α_1 and λ for various M, number of unknown parameters

Using the above distribution and the optimal solution in (82) we solve for the optimal solution to $\alpha_2(w_1)$ at the second stage. We already know that the minimum occurs at $\alpha_1 \approx 0.7421$ for $M = 1$. We solve for the optimal solution and find the gain in MSE for various values of M. The optimal values of α_1 and λ' for varying values of M is shown in Table 2. In particular, the performance of the two-step sequential design are plotted versus varying values of α_1 theoretically (solid) and via simulations (dashed dotted) for $M = 2$ and $M = 3$ case in Fig. 13 and 14, respectively. Figure 15 plots the optimal reduction for the two-step design for increasing values of M theoretically (solid) and via simulations (dashed dotted). It is interesting to note that the reduction in MSE decreases as the number of parameters increases. In fact, as the number of unknown parameters M goes to infinity, the ratio tends to 1. We will prove this fact in the following sections.

8.2.2 Suboptimal Solution

As in the case of the single parameter case, we explore the performance gain for a suboptimal solution of the form,

$$\alpha_2(w_1) = \alpha_2 I(w_1 \geq \rho), \tag{97}$$

where α_2, ρ are chosen to satisfy the average energy constraint which can be written as

$$\alpha_1^2 + \alpha_2^2 \left(1 - F_\chi(w_1)\right) \leq 1. \tag{98}$$

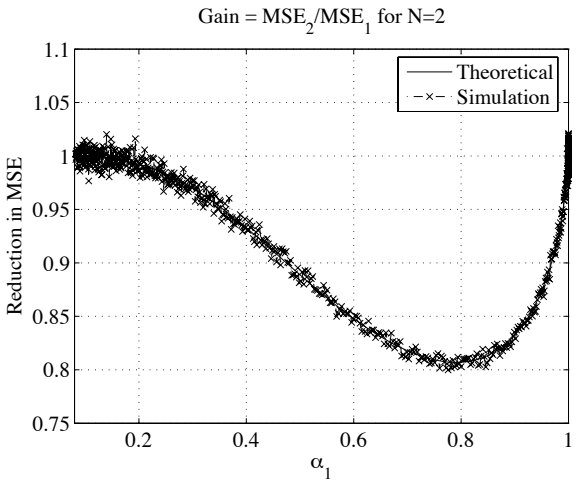


Figure 13: Plot of gain in two-step sequential design versus α_1 for $M = 2$ through theory and simulations

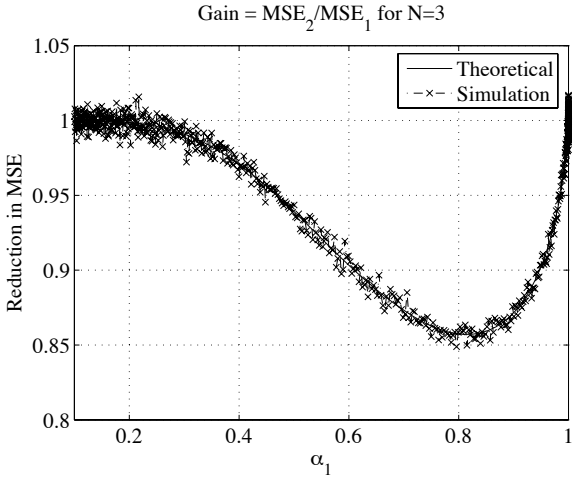


Figure 14: Plot of gain in two-step sequential design versus α_1 for $M = 3$ through theory and simulations

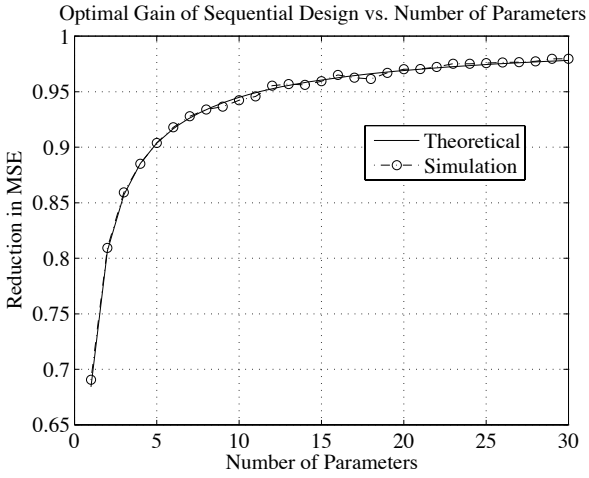


Figure 15: Plot of gain in two-step sequential design versus number of parameters M through theory and simulations

This solution is motivated from the suboptimal solution presented in Section 4 for the scalar parameter case. Simplifying the energy constraint, we have

$$\alpha_2^2 = \frac{1-\alpha_1^2}{1-F_\chi(w_1)}, \tag{99}$$

where $F_\chi(w_1) = \mathrm{P}(w_1 \leq \rho)$ is the cumulative distribution function of the chi-square mixture. Substituting (97) in (73) yields,

$$\begin{aligned}
\mathrm{tr}(\mathrm{MSE}^{(2)}) &= \mathrm{tr}(\mathrm{MSE}^{(1)})\mathrm{E}_{w_1}\left[\frac{\alpha_1^2}{(\alpha_1^2+\alpha_2^2)^2} w_1 I(w_1 \geq \rho)\right. \\
&\quad \left. + \frac{\alpha_2^2}{(\alpha_1^2+\alpha_2^2)^2} I(w_1 \geq \rho) + \frac{1}{\alpha_1^2} w_1 I(w_1 < \rho)\right] \qquad (100) \\
&= \mathrm{tr}(\mathrm{MSE}^{(1)}) \left(\frac{1}{\alpha_1^2} + \left\{ \frac{\alpha_1^2}{(\alpha_1^2+\alpha_2^2)^2} - \frac{1}{\alpha_1^2} \right\} \mathrm{E}_{w_1}\left[w_1 I(w_1 \geq \rho)\right] \right. \\
&\quad \left. + \frac{\alpha_2^2}{(\alpha_1^2+\alpha_2^2)^2} \left(1 - F_\chi(w_1)\right) \right), \qquad (101)
\end{aligned}$$

where α_2 satisfies (99). We minimize the expression in (101) over $0 \leq \alpha_1 \leq 1$ and $\rho \in \mathcal{R}^+$. Figure 16 plots the gain in MSE as a function of the number of unknown parameters for the suboptimal solution through simulations along with the optimal gain. As is the case for a single parameter case, the best possible suboptimal design gives us an improvement of approximately 0.71.

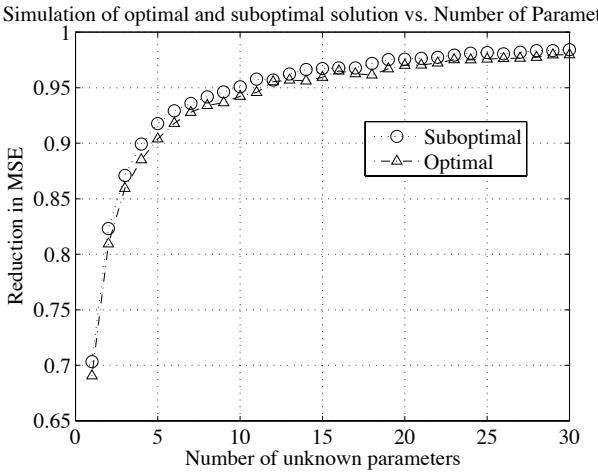


Figure 16: Plot of gains obtained in a suboptimal design versus number of unknown parameters to be estimated through simulations

Also, it is worthwhile to note that the suboptimal solution provides near optimal performance. This is essentially due to the fact that the suboptimal solution is in accordance with the structure of optimal solution.

Note: The general two-step design procedure is given in terms of w_1 or in terms of $\{\widehat{\mathbf{n}}_{1,i}\}_{i=1}^{N}$ as in (74). Since

$$\widehat{\mathbf{n}}_1 \;=\; \mathbf{Q}^H\mathbf{R}_{\mathbf{n}}\mathbf{n}_1 = \mathbf{Q}^H\mathbf{R}_{\mathbf{n}}\left(\mathbf{y}_1 - \sqrt{E_0}\alpha_1\widetilde{\mathbf{X}}\boldsymbol{\theta}\right), \tag{102}$$

it follows that the unknown $\boldsymbol{\theta}$ can be replaced by a guess $\boldsymbol{\theta}_g$ and a $K\times 2$-step procedure will yield the desired performance as in the single parameter case.

8.2.3 Asymptotic Behaviour of Optimal Design

Asymptotic distribution of w_1: When $d_1 = \ldots = d_N$, w_1 is a gamma distribution and it follows that the random variable w_1 asymptotically behaves as [49]

$$w_1 \sim \frac{1}{N}\chi_N \rightarrow \mathcal{N}(1, \frac{2}{N}) \quad as\ N \rightarrow \infty \tag{103}$$

Theorem 8.4. *$tr(MSE_2) \rightarrow tr(MSE_1)$ as $M \rightarrow \infty$, i.e., There is no gain in a sequential design procedure when the number of unknown parameters M goes to infinity.*

Proof. Since $N \geq M$, as $M \to \infty$, we have $w_1 \to \mathcal{N}(1, 2/N)$. It follows that asymptotically, the optimal design at the second step should concentrate all the energy at $w_1 = 1$, i.e., $\alpha_2(w_1) = \alpha_2\delta(w_1 - 1)$ and the corresponding energy constraint is $\alpha_1^2 + \alpha_2^2 \leq 1$. Hence the minimum MSE at the second step asymptotically is given by

$$\begin{aligned}
\mathrm{tr}(\mathrm{MSE}_2) &= \mathrm{tr}(\mathrm{MSE}_1)\mathrm{E}_{w_1}\left[\frac{\alpha_1^2}{(\alpha_1^2+\alpha_2^2(w_1))^2}w_1 + \frac{\alpha_2^2(w_1)}{(\alpha_1^2+\alpha_2^2(w_1))^2}\right] \\
&= \lim_{M\to\infty} \mathrm{tr}(\mathrm{MSE}_1)\mathrm{E}_{w_1}\left[\frac{\alpha_1^2}{(\alpha_1^2+\alpha_2^2\delta(w_1-1))^2}w_1 + \frac{\alpha_2^2\delta(w_1-1)}{(\alpha_1^2+\alpha_2^2\delta(w_1-1))^2}\right] \\
&= \lim_{M\to\infty} \mathrm{tr}(\mathrm{MSE}_1)\frac{1}{\alpha_1^2+\alpha_2^2} \\
&= \mathrm{tr}(\mathrm{MSE}_1)
\end{aligned}$$

□

8.2.4 Waveform and energy design for trace criterion

The problem of multiple parameter estimation is more complicated than estimation of a single parameter for the following reason. We showed in Section 2.1 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

$$\mathbf{H}(\mathbf{x}) = \begin{bmatrix} x_1 & x_2 \\ 0 & x_2 \end{bmatrix}, \tag{104}$$

$\mathbf{x} = [x_1\ x_2]^T$, $\mathbf{y} = [y_1\ y_2]^T$, $\mathbf{n} = [n_1\ n_2]^T \sim \mathcal{CN}(\mathbf{0}, \mathbf{R_n})$, and $\mathbf{R_n} = \sigma^2\mathbf{I}$. 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 $\|\mathbf{x}\|^2 \leq E_0 = 1$, 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,

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

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 7 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 in (47). In other words, the MSE gains of the N-step procedure carry over to the vector parameter case as well.

9 Applications of sequential estimation

9.1 MIMO Channel Estimation

It has been shown that multiple-input and multiple-output systems (MIMO) greatly increase the capacity of wireless systems [50–52] and hence MIMO has become an active area of research over the last decade [53, 54]. 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 [55]. 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 [56], least squares training [57], minimum mean squared estimation [58]. 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.

9.1.1 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, 59]

$$\mathbf{R} = \mathbf{\Theta X} + \mathbf{M}, \tag{105}$$

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., $\|\mathbf{X}\|_{\mathrm{F}}^2 = P_0$, $\|\cdot\|_F$ indicates Frobenius norm ($\|\mathbf{X}\|_{\mathrm{F}} = \sqrt{\mathrm{tr}(\mathbf{X}^H\mathbf{X})}$) and σ^2 denote the variance of receiver noise. Though $\mathbf{\Theta}$ is random, we estimate $\mathbf{\Theta}$ for a particular realization corresponding to the block of received data. The task of channel estimation is to recover the channel matrix $\mathbf{\Theta}$ 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

$$\widehat{\mathbf{\Theta}}_{\mathrm{LS}} = \mathbf{R}\mathbf{X}^H(\mathbf{X}\mathbf{X}^H)^{-1} \tag{106}$$

$$\mathrm{MSE}_{\mathrm{LS}} = \frac{\sigma^2 t^2 r}{P_0}. \tag{107}$$

Assuming co-located transmitter and receiver arrays [60, 61] 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 $\mathbf{R}_i = \mathbf{\Theta X}_i + \mathbf{M}_i,\ i = 1, 2, \ldots, N$, can be rewritten in the following form:

$$\mathbf{y}_i = \mathbf{H}(\mathbf{X}_i)\boldsymbol{\theta} + \mathbf{n}_i, \quad i = 1, 2, \ldots, N, \tag{108}$$

where $\mathbf{y}_i = \mathrm{vec}(\mathbf{R}_i), \boldsymbol{\theta} = \mathrm{vec}(\mathbf{\Theta}), \mathbf{n}_i = \mathrm{vec}(\mathbf{M}_i), \mathrm{vec}(\cdot)$ 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 (2). 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 $\mathrm{MSE}_{\mathrm{LS}}^{(N)} = \sigma^2 t^2 r / P_0$, where P_0 is the total power used in the N steps, i.e., $\sum_{i=1}^{N} \|\mathbf{X}_i\|_{\mathrm{F}}^2 \leq P_0$. 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 (108) along with the average power constraint $\mathrm{E}\left[\sum_i \|\mathbf{X}_i\|_{\mathrm{F}}^2\right] \leq P_0$ can benefit from adaptive energy allocation designs in Sections 7 and 8.2, 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 $\mathrm{MSE}^{(N)} \approx \mathcal{G}_N \sigma^2 t^2 r / P_0$,

which we have shown to be at least 5dB (in 50 steps) better than any non-adaptive strategy.

9.2 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 [62], foliage penetrating radar, non-destructive testing [63], and active audio. The goal in imaging is to detect and image small scatterers in a known background medium. A recent approach [64] 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 [65]. 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.

9.2.1 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

$$\mathbf{a}_i = \left[\left(\frac{\exp(-j\omega/c \|\mathbf{r}_k^a - \mathbf{r}_i^v\|)}{\|\mathbf{r}_k^a - \mathbf{r}_i^v\|} \right)_{k=1\ldots N} \right]^T, \tag{109}$$

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., [66]. 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 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,

$$\mathbf{y}_i = \mathbf{A}\text{diag}(\boldsymbol{\theta})\mathbf{A}^T\mathbf{x}_i + \mathbf{n}_i = \mathbf{H}(\mathbf{x}_i)\boldsymbol{\theta} + \mathbf{n}_i, \ i = 1, 2, \ldots, N, \tag{110}$$

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 complex normal random vectors with zero mean and a covariance matrix $\sigma^2\mathbf{I}$. 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 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 8.2, 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 8.1 can be used to achieve improvement in performance.

10 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.

11 Appendix: Proof of equivalence

Denote the set of design parameters as $\mathbf{X} = \{\mathbf{x}_i\,(\mathbf{y}_1, \cdots, \mathbf{y}_{i-1})\}_{i=1}^{N}$. Let

$$\mathbf{X}^{+} = \arg\min_{\mathbf{X}} \ \mathrm{MSE}^{(N)}(\mathbf{X})\mathrm{SNR}^{(N)}(\mathbf{X}) \tag{111}$$

$$\mathbf{X}^{*} = \arg\min_{\mathbf{X}} \ \mathrm{MSE}^{(N)}(\mathbf{X}) \ \text{s.t} \quad \mathrm{SNR}^{(N)}(\mathbf{X}) \leq \mathrm{SNR}_0. \tag{112}$$

Lemma 11.1. *For any $\beta \in \mathbb{R}$, $\beta\mathbf{X}^{+}$ is also a minimizer of the minimization criterion in* (111)*, where $\beta\mathbf{X} = \{\beta\mathbf{x}_i\,(\mathbf{y}_1, \cdots, \mathbf{y}_{i-1})\}_{i=1}^{N}$.*

Proof. From the energy definition in (6), the SNR definition in (10), and the property $\|\beta\mathbf{x}\| = \beta\|\mathbf{x}\|$, we obtain $\mathrm{SNR}^{(N)}(\beta\mathbf{X}^{+}) = \beta^2\mathrm{SNR}^{(N)}(\mathbf{X}^{+})$. Using the scaling property of the linearity of $\mathbf{h}_1(\cdot)$, $\mathbf{h}_1(\beta\mathbf{x}) = \beta\ \mathbf{h}_1(\mathbf{x})$ in (5), we have $\mathrm{MSE}^{(N)}(\beta\mathbf{X}^{+}) = \frac{1}{\beta^2}\mathrm{MSE}^{(N)}(\mathbf{X}^{+})$. Hence $\mathrm{MSE}^{(N)}(\beta\mathbf{X}^{+})\mathrm{SNR}^{(N)}(\beta\mathbf{X}^{+}) = \mathrm{MSE}^{(N)}(\mathbf{X}^{+})\mathrm{SNR}^{(N)}(\mathbf{X}^{+})$, which is the minimum value of the criterion in (111). □

Since $\mathbf{X}^{*}$ minimizes the RHS of (112), we have

$$\mathrm{MSE}^{(N)}(\mathbf{X}^{*}) \leq \mathrm{MSE}^{(N)}(\beta\mathbf{X}^{+}), \tag{113}$$

where β satisfies $\mathrm{SNR}^{(N)}(\beta\mathbf{X}^{+}) \leq \mathrm{SNR}_0$. Similarly, from Lemma 11.1, we obtain

$$\mathrm{MSE}^{(N)}(\beta\mathbf{X}^{+})\mathrm{SNR}^{(N)}(\beta\mathbf{X}^{+}) \leq \mathrm{MSE}^{(N)}(\mathbf{X}^{*})\mathrm{SNR}^{(N)}(\mathbf{X}^{*}). \tag{114}$$

Multiplying (113) by $\mathrm{SNR}^{(N)}(\beta\mathbf{X}^{+})$ and combining it with (114) yields

$$\mathrm{MSE}^{(N)}(\mathbf{X}^{*})\mathrm{SNR}^{(N)}(\beta\mathbf{X}^{+}) \leq \mathrm{MSE}^{(N)}(\beta\mathbf{X}^{+})\mathrm{SNR}^{(N)}(\beta\mathbf{X}^{+}) \leq \mathrm{MSE}^{(N)}(\mathbf{X}^{*})\mathrm{SNR}^{(N)}(\mathbf{X}^{*}), \tag{115}$$

for $|\beta| \leq \sqrt{\frac{\mathrm{SNR}_0}{\mathrm{SNR}^{(N)}(\mathbf{X}^{+})}}$. Choosing $\beta = \sqrt{\frac{\mathrm{SNR}^{(N)}(\mathbf{X}^{*})}{\mathrm{SNR}^{(N)}(\mathbf{X}^{+})}}$ satisfies this constraint and $\mathrm{SNR}^{(N)}(\beta\mathbf{X}^{+}) = \mathrm{SNR}^{(N)}(\mathbf{X}^{*}) \leq \mathrm{SNR}_0$. Replacing $\mathrm{SNR}^{(N)}(\beta\mathbf{X}^{+})$ with $\mathrm{SNR}^{(N)}(\mathbf{X}^{*})$ in (115), we obtain

$$\mathrm{MSE}^{(N)}(\mathbf{X}^{*})\mathrm{SNR}^{(N)}(\mathbf{X}^{*}) \leq \mathrm{MSE}^{(N)}(\beta\mathbf{X}^{+})\mathrm{SNR}^{(N)}(\mathbf{X}^{*}) \leq \mathrm{MSE}^{(N)}(\mathbf{X}^{*})\mathrm{SNR}^{(N)}(\mathbf{X}^{*}), \tag{116}$$

It follows that $\mathrm{MSE}^{(N)}(\mathbf{X}^{*}) = \mathrm{MSE}^{(N)}(\beta\mathbf{X}^{+})$, i.e., $\beta\mathbf{X}^{+}$ with $\beta = \sqrt{\frac{\mathrm{SNR}^{(N)}(\mathbf{X}^{*})}{\mathrm{SNR}^{(N)}(\mathbf{X}^{+})}}$ is the minimizer to the constrained minimization problem in (112). Furthermore,

Lemma 11.2. *$SNR^{(N)}(\mathbf{X}^*) = SNR_0$.*

Proof. By contradiction: If $\mathrm{SNR}^{(N)}(\mathbf{X}^*) < \mathrm{SNR}_0$, let $\mathrm{SNR}^{(N)}(\mathbf{X}^*) = \frac{1}{\beta}\mathrm{SNR}_0$ for some $\beta > 1$. Then by using the property $\|\alpha\mathbf{x}\| = \alpha\|\mathbf{x}\|$ in (6) and (10), we have $\mathrm{SNR}^{(N)}(\sqrt{\beta}\mathbf{X}^*) = \mathrm{SNR}_0$. Using linearity of $\mathbf{h}_1(\cdot)$ in (5), we obtain $\mathrm{MSE}^{(N)}(\sqrt{\beta}\mathbf{X}^*) = \frac{1}{\beta}\mathrm{MSE}^{(N)}(\mathbf{X}^*) < \mathrm{MSE}^{(N)}(\mathbf{X}^*)$. It follows that $\sqrt{\beta}\mathbf{X}^*$ satisfies the constraint and achieves a lower MSE than $\mathbf{X}^*$ which contradicts the fact that $\mathbf{X}^*$ is the minimum. □

12 Appendix: Solution to Problem 6.1

12.0.2 Properties of $\eta(z)$ in (40)

We list some of the properties of $\eta(z)$ which we will use to prove our results.

Proposition 12.1. *$\eta(z)$ achieves two-step minimum, i.e., $\eta(z)_{|z=0} = \eta^*$.*

Proposition 12.2. *$\eta(z)$ is an even function of z, i.e., $\eta(z) = \eta(|z|)$.*

Proof. $\eta(z)$ in (40) depends on z only through expected values of the form $\mathrm{E}_{\tilde{n}_1}[f(|\tilde{n}_1 + z|)]$ for some continuous function f. Thus $\mathrm{E}_{\tilde{n}_1}[f(|\tilde{n}_1 + z|)] = \mathrm{E}_{\tilde{n}_1}\left[f\left(\left|e^{-\angle z}\tilde{n}_1 + |z|\right|\right)\right] = \mathrm{E}_{\tilde{m}_1}[f(|\tilde{m}_1 + |z||)]$, where $\tilde{m}_1 = e^{-\angle z}\tilde{n}_1$ is another complex Gaussian random variable with zero mean and unit variance. □

Proposition 12.3. *Optimal two-step minimum is achieved uniquely: $\eta(z) > \eta(0)\ \forall z \in \mathbb{C} - \{0\}$. Therefore, $MSE^{(2)}(z) \times SNR^{(2)}(z) = \eta(z)$ achieves a global minimum at $z = 0$, or $\theta_g = \theta_1$.*

Proof. By contradiction. If there exists z^+ such that $\eta(z^+) < \eta(0)$, then the design parameters α_1^* and $\alpha_2^*(\tilde{n}_1 + z^+)$ will yield a $\mathrm{MSE}^{(2)} \times \mathrm{SNR}^{(2)} < \eta(0)$ which contradicts the fact that α_1^* and α_2^* achieves minimal $\eta(0)$. □

Proposition 12.4. *Continuity of $\eta(z)$: For any $\epsilon > 0, \exists\ \delta > 0$ such that for any $0 \le z \le \delta$, $\eta^* \le \eta(z) \le (1+\epsilon)\eta^*$, i.e., the optimal performance $MSE^{(2)} \times SNR^{(2)} = \eta^*$ can be approached within ϵ for any θ_g lying in the sphere $0 \le |\theta_g - \theta_1| \le \frac{\delta}{\alpha_1\sqrt{SNR_0}}$. Note that the sphere size increases as SNR decreases.*

Proof. It follows from the fact that the functions $\mathrm{E}[\cdot], \alpha_2^*(\cdot), f(\tilde{n}_1)$ are continuous. If the solution to $\alpha_2^*(\cdot)$ is the suboptimal thresholding function of the form (25) then the $\mathrm{MSE}^{(2)}(z)$ and $\mathrm{SNR}^{(2)}(z)$ are integrals of the probability density function of an independent Gaussian random variable over ellipsoids whose center is given through z and hence are still continuous functions in z. □

12.1 The $N\times$ two-step procedure

In this section, we take advantage of Proposition 12.4 to prove our result, i.e., the fact that the suboptimal solution presented in Section 6.2 can approach the optimal two-step solution in Section 3 when θ_g lies within a sphere centered at θ_1 with radius which increases as SNR decreases.

Consider N independent two-step experiments described below. The observations from the k^{th} experiment are

$$\begin{aligned} \mathbf{y}_1^k &= \mathbf{h}_1(\mathbf{x}_1^k)\theta_1 + \mathbf{n}_1^k && (117)\\ \mathbf{y}_2^k &= \mathbf{h}_1(\mathbf{x}_2^k)\theta_1 + \mathbf{n}_2^k,\ k = 1, 2, \ldots, N, && (118) \end{aligned}$$

where $\mathbf{n}_1^k$ and $\mathbf{n}_2^k$ are i.i.d $\mathcal{CN}(\mathbf{0}, \sigma^2\mathbf{I})$. The input design vectors for the N experiments are given by

$$\begin{aligned} \mathbf{x}_1^k &= \sqrt{\frac{E_0}{N(1+\epsilon)}}\alpha_1^*\mathbf{v}_{\text{m}} && (119)\\ \mathbf{x}_2^k &= \sqrt{\frac{E_0}{N(1+\epsilon)}}\alpha_2^*\left(\left|\frac{\mathbf{h}_1(\mathbf{v}_{\text{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\text{m}})\|}\frac{(\mathbf{y}_1^k - \mathbf{h}_1(\mathbf{x}_1^k)\theta_g)}{\sigma}\right|\right)\mathbf{v}_{\text{m}}\\ &= \sqrt{\frac{E_0}{N(1+\epsilon)}}\alpha_2^*\left(\left|\tilde{n}_1^k(\mathbf{y}_1^k;\theta_1) + z^{'}\right|\right)\mathbf{v}_{\text{m}},\ k = 1, 2, \ldots, N, && (120) \end{aligned}$$

where

$$z^{'} = \frac{z}{\sqrt{N(1+\epsilon)}} = \sqrt{\frac{\text{SNR}_0}{N(1+\epsilon)}}\alpha_1^*(\theta_1 - \theta_g), \qquad (121)$$

$$\begin{aligned} \tilde{n}_1^k(\mathbf{y}_1^k;\theta_1) &= \frac{\mathbf{h}_1(\mathbf{v}_{\text{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\text{m}})\|}\left(\frac{\mathbf{y}_1^k - \mathbf{h}_1(\mathbf{x}_1^k)\theta_1}{\sigma}\right)\\ &= \frac{\mathbf{h}_1(\mathbf{v}_{\text{m}})^H}{\|\mathbf{h}_1(\mathbf{v}_{\text{m}})\|}\frac{\mathbf{n}_1^k}{\sigma},\ k = 1, 2, \ldots, N, && (122) \end{aligned}$$

and $\epsilon > 0$. The SNR in each experiment is

$$\begin{aligned} \text{SNR}^{(2),k}(z^{'}) &= \frac{\lambda_{\text{m}}(\boldsymbol{H}_1)}{\sigma^2}\text{E}\left[\|\mathbf{x}_1^k\|^2 + \|\mathbf{x}_2^k\|^2\right]\\ &= \frac{\text{SNR}_0}{N(1+\epsilon)}\left(\alpha_1^{*2} + \text{E}\left[\alpha_2^{*2}\left(\left|\tilde{n}_1^k + z^{'}\right|\right)\right]\right). && (123) \end{aligned}$$

Then the overall SNR in the $2N$ experiments (N two-step procedures) is given by

$$\begin{aligned}\mathrm{SNR}^{(2N)}(z^{'}) &= \sum_{k=1}^{N}\mathrm{SNR}^{(2),k}(z^{'}) \\ &= N\ \mathrm{SNR}^{(2),1}(z^{'}) \\ &= \frac{\mathrm{SNR}_0}{(1+\epsilon)}\left(\alpha_1^{*2}+\mathrm{E}\left[\alpha_2^{*2}\left(\left|\tilde{n}_1^1+z^{'}\right|\right)\right]\right), \end{aligned} \tag{124}$$

since $\{\tilde{n}_1^k\}_{k=1}^N$ are i.i.d $\mathcal{CN}(0,1)$ and hence the expected value of $\alpha_2^{*2}(\cdot)$ is identical and independent of k. The ML estimator for the k^{th} two-step process, $\hat{\theta}_1^{(2),k}$ is given by

$$\hat{\theta}_1^{(2),k} = \frac{\{\mathbf{h}_1(\mathbf{x}_1^k)\}^H\mathbf{y}_1^k+\{\mathbf{h}_1(\mathbf{x}_2^k)\}^H\mathbf{y}_2^k}{\|\mathbf{h}_1(\mathbf{x}_1^k)\|^2+\|\mathbf{h}_1(\mathbf{x}_2^k)\|^2}. \tag{125}$$

Since the N experiments are independent, the estimators $\{\hat{\theta}_1^{(2),k}\}_{k=1}^N$ are i.i.d random variables. The ML estimator for the $N\times 2$-step procedure is

$$\hat{\theta}_1^{(2N)} = \frac{1}{N}\sum_{k=1}^{N}\hat{\theta}_1^{(2),k} \tag{126}$$

and the corresponding MSE is

$$\mathrm{MSE}^{(2N)}(z^{'}) = \mathrm{E}\left[\left|\theta_1-\hat{\theta}_1^{(2N)}\right|^2\right] = \frac{1}{N^2}\sum_{k=1}^{N}\mathrm{MSE}^{(2),k}(z^{'}) = \frac{1}{N}\mathrm{MSE}^{(2),1}(z^{'}), \tag{127}$$

where $\mathrm{MSE}^{(2),k}(z^{'}) = \mathrm{E}\left[\left|\theta_1-\hat{\theta}_1^{(2),k}\right|^2\right]$ are identical as the N two-step experiments are independent. Further, the input designs for each 2-step experiment given by (119) and (120) are the same as the suboptimal input designs in (37) and (38) with energy $E_0/(N(1+\epsilon))$, and using a similar derivation to (40), we obtain

$$\mathrm{MSE}^{(2),k}(z^{'}) = \frac{N(1+\epsilon)}{\mathrm{SNR}_0}\mathrm{E}\left[\frac{\alpha_1^{*2}\left|\tilde{n}_1^k+z^{'}\right|^2+\alpha_2^{*2}\left(\left|\tilde{n}_1^k+z^{'}\right|\right)}{\left(\alpha_1^{*2}+\alpha_2^{*2}\left(\left|\tilde{n}_1^k+z^{'}\right|\right)\right)^2}\right]. \tag{128}$$

Lemma 12.5. *Given $\epsilon>0$, $\theta_1,\theta_g\in[\theta_{\min},\theta_{\max}]$, $\theta_{\min},\theta_{\max}\in\mathbb{R}$, $\exists N_0$ such that $\forall N\geq N_0$, $\frac{SNR_0}{(1+\epsilon)}\leq SNR^{(2N)}(z^{'})\leq SNR_0$.*

Proof. $\frac{\mathrm{SNR}_0}{(1+\epsilon)} \leq \mathrm{SNR}^{(2N)}(z^{'}), \forall N \in \mathbb{N}$ follows from the fact $\mathrm{SNR}^{(2N)}(z^{'})$ is an even function of z and achieves minimum at $z^{'} = 0$. (12.5.1)

Since $\mathrm{E}[\cdot]$ and $\alpha_2^*(\cdot)$ are continuous functions, $\mathrm{SNR}^{(2N)}(z^{'})$ is continuous everywhere and hence at $z^{'} = 0$. Thus for every $\zeta > 0$, $\exists \delta > 0$ such that $|z^{'} - 0| < \delta$ implies $|\mathrm{SNR}^{(2N)}(z^{'}) - \frac{\mathrm{SNR}_0}{(1+\epsilon)}| < \zeta$. Choose $\zeta = \frac{\epsilon\ \mathrm{SNR}_0}{(1+\epsilon)}$, we have $\mathrm{SNR}^{(2N)}(z^{'}) \leq \mathrm{SNR}_0$. (12.5.2)

Choose $N_0 = \left\lceil \frac{\alpha_1^{*2}\lambda_{\mathrm{m}}(\boldsymbol{H}_1)E_0|\theta_{\max}-\theta_{\min}|^2}{\sigma^2\delta^2(1+\epsilon)} \right\rceil$, from (12.5.1) and (12.5.2), the result follows. Q.E.D. □

Theorem 12.6. *Given $\Delta > 0$, $\theta_1, \theta_g \in [\theta_{\min}, \theta_{\max}]$, $\exists N_0$ such that $\forall N \geq N_0$ we have $\frac{\eta^*}{SNR_0} \leq MSE^{(2N)}(z^{'}) \leq \frac{(1+\Delta)\eta^*}{SNR_0}$.*

In other words, this theorem states that we can asymptotically achieve the performance of the optimal two-step estimator η^*/SNR_0 using the $N\times$ two-step procedure when θ_1 is bounded.

Proof. Comparing the product of expressions in (128) and (124) to the expression for $\eta(\cdot)$ in (40) and using Proposition 12.2, we have for a single two-step procedure that

$$\begin{aligned} \mathrm{MSE}^{(2),k}(z^{'}) \times \mathrm{SNR}^{(2),k}(z^{'}) &= \eta\left(z^{'}\right) \\ &= \eta\left(\alpha_1^*|\theta_1 - \theta_g|\sqrt{\frac{\mathrm{SNR}_0}{N(1+\epsilon)}}\right). \end{aligned} \quad (129)$$

Since $\mathrm{MSE}^{(2N)} = \frac{1}{N}\mathrm{MSE}^{(2),k}$ and $\mathrm{SNR}^{(2N)} = N\mathrm{SNR}^{(2),k}$, the total MSE satisfies

$$\begin{aligned} \mathrm{MSE}^{(2N)}(z^{'}) \times \mathrm{SNR}^{(2N)}(z^{'}) &= \frac{1}{N}\mathrm{MSE}^{(2),k}(z^{'}) \times N\mathrm{SNR}^{(2),k}(z^{'}) \\ &= \eta\left(\alpha_1^*|\theta_1 - \theta_g|\sqrt{\frac{\mathrm{SNR}_0}{N(1+\epsilon)}}\right). \end{aligned} \quad (130)$$

Using Proposition 12.3, $\mathrm{MSE}^{(2N)}(z^{'}) \times \mathrm{SNR}^{(2N)}(z^{'}) \geq \eta^*$. From RHS of *Lemma* 12.5, it follows that

$$\mathrm{MSE}^{(2N)}(z^{'}) \geq \frac{\eta^*}{\mathrm{SNR}^{(2N)}(z^{'})} \geq \frac{\eta^*}{\mathrm{SNR}_0}. \quad (131)$$

Since $\eta(\cdot)$ is continuous everywhere and at $z^{'} = 0$, it follows that for every $\mu^{'} > 0$, $\exists \delta^{'} > 0$ such that $|z^{'} - 0| \leq \delta^{'}$ implies $|\eta(z^{'}) - \eta^*| \leq \mu^{'}$. Choose $\mu^{'} =$

$\eta^* \epsilon^{'}$, we obtain $\eta(z^{'}) \leq (1+\epsilon^{'})\eta^*$. Thus for $N_0 = \left\lceil \frac{\alpha_1^{*2}\lambda_{\rm m}(\boldsymbol{H}_1)E_0|\theta_{\max}-\theta_{\min}|^2}{\sigma^2(1+\epsilon)} \right\rceil \max\left(\frac{1}{\delta}, \frac{1}{\delta'}\right)^2$, we have

$$\begin{aligned} \mathrm{MSE}^{(2N)}(z^{'}) &\leq \frac{(1+\epsilon^{'})\eta^*}{\mathrm{SNR}^{(2N)}(z^{'})} \\ &\leq \frac{(1+\epsilon^{'})(1+\epsilon)\eta^*}{\mathrm{SNR}_0} \quad \text{from } \textit{Lemma } 12.5 \\ &\leq \frac{(1+\Delta)\eta^*}{\mathrm{SNR}_0}, \end{aligned} \tag{132}$$

where $\epsilon^{'} = \left(\frac{\Delta-\epsilon}{1+\epsilon}\right)$ and $0 < \epsilon < \Delta$. From (131) and (132), we have the result. □

13 Appendix: Derivation of the N-step procedure

The design of the N-step procedure given by (42) can be written as

$$\alpha_i = A_i \prod_{s=1}^{i-1} I\left(|w_s|^2 \geq \rho_{s+1}\right),\ 1 \leq i \leq N, \tag{133}$$

where

$$w_s = \frac{\sum_{j=1}^{s} A_j \tilde{n}_j}{\sqrt{\sum_{j=1}^{s} |A_j|^2}}. \tag{134}$$

Then $\mathbf{w} = [w_1, \ldots, w_N]^T$ is a zero mean complex normal vector. Define the sets $U_i = \{|w_i|^2 \geq \rho_{i+1}\}, 1 \leq i \leq N$. For the set of events $\{U_i\}_{i=1}^{N-1}$, the set of events $\{D_i = \cap_{k=1}^{i-1} U_k \cap U_i^c,\ 1 \leq i \leq N-1, \cap_{k=1}^{N-1} U_k,\ i = N\}$ are disjoint and satisfy $I(\mathbf{w} \in \{D_i\}_{i=1}^{N}) = 1,\ \forall \mathbf{w} \in \mathbb{C}^N$. Hence

$$I\left(I(|w_1|^2 < \rho_2) + \sum_{i=2}^{N-1} I(|w_i|^2 < \rho_{i+1}) \prod_{s=1}^{i-1} I(|w_s|^2 \geq \rho_{s+1}) + \prod_{s=1}^{N-1} I(|w_s|^2 > \rho_{s+1})\right) = 1.$$

Substituting this expression inside the expectation for the $\mathrm{MSE}^{(N)}$ in (5), we obtain

$$\begin{aligned} \mathrm{MSE}^{(N)} = \mathrm{E}\Bigg[\frac{|\sum_{i=1}^{N} \mathbf{h}_1(\mathbf{x}_i)^H \mathbf{n}_i|^2}{|\sum_{i=1}^{N} \|\mathbf{h}_1(\mathbf{x}_i)\|^2|^2} \Bigg(& I(|w_1|^2 < \rho_2) + \\ & \sum_{i=2}^{N-1} I(|w_i|^2 < \rho_{i+1}) \prod_{s=1}^{i-1} I(|w_s|^2 \geq \rho_{s+1}) + \prod_{s=1}^{N-1} I(|w_s|^2 > \rho_{s+1}) \Bigg) \Bigg] \end{aligned}$$

$$
\begin{aligned}
&= \frac{\sigma^2}{\lambda_{\rm m}(\boldsymbol{H}_1)}\Bigg(\frac{1}{|A_1|^2}{\rm E}\left[|w_1|^2 I(|w_1|^2<\rho_2)\right] \\
&\quad + \sum_{i=2}^{N-1}\frac{1}{\sum_{j=1}^{i}|A_j|^2}{\rm E}\left[|w_i|^2 I(|w_i|^2<\rho_{i+1})\prod_{s=1}^{i-1} I(|w_s|^2>\rho_{s+1})\right] \\
&\quad + \frac{}{\sum_{j=1}^{N}|A_j|^2}{\rm E}\left[|w_N|^2\prod_{s=1}^{N-1} I(|w_s|^2>\rho_{s+1})\right]\Bigg) \\
{\rm MSE}^{(N)} &= \frac{\sigma^2}{\lambda_{\rm m}(\boldsymbol{H}_1)}\left\{\sum_{i=1}^{N-1}\frac{T_i}{Q_i}+\frac{\tilde{T}_N}{Q_N}\right\}, \qquad (135)
\end{aligned}
$$

where

$$
\begin{aligned}
Q_i &= \sum_{s=1}^{i}|A_s|^2, \qquad (136)\\
T_1 &= {\rm E}\left[|w_1|^2 I(|w_1|^2<\rho_2)\right] = 1-(1+\rho_2)e^{-\rho_2}, \\
T_i &= {\rm E}\left[|w_i|^2 I(|w_i|^2<\rho_{i+1})\prod_{s=1}^{i-1} I(|w_s|^2>\rho_{s+1})\right], \qquad (137)\\
\tilde{T}_i &= {\rm E}\left[|w_i|^2\prod_{s=1}^{i-1} I(|w_s|^2>\rho_{s+1})\right].
\end{aligned}
$$

The SNR of this N-step process is given by

$$
\begin{aligned}
{\rm SNR}^{(N)} &= \frac{\lambda_{\rm m}(\boldsymbol{H}_1)}{\sigma^2}{\rm E}\left[\sum_{i=1}^{N}\|\mathbf{x}_i\|^2\right] \\
&= \frac{\lambda_{\rm m}(\boldsymbol{H}_1)}{\sigma^2}\Bigg\{A_1^2{\rm E}\left[I(|w_1|^2<\rho_2)\right] \\
&\quad + \sum_{i=2}^{N-1}\left(\sum_{j=1}^{i-1}|A_j|^2\right){\rm E}\left[I(|w_i|^2<\rho_{i+1})\prod_{s=1}^{i-1} I(|w_s|^2\geq\rho_{s+1})\right] \\
&\quad + \left(\sum_{j=1}^{N}|A_j|^2{\rm E}\left[\prod_{s=1}^{N-1} I(|w_s|^2\geq\rho_{s+1})\right]\right)\Bigg\} \\
&= \frac{\lambda_{\rm m}(\boldsymbol{H}_1)}{\sigma^2}\left\{\sum_{i=1}^{N-1}Q_iP_i+Q_N\tilde{P}_N\right\}, \qquad (138)
\end{aligned}
$$

where

$$\begin{aligned} P_1 &= \mathrm{E}\left[I(|w_1|^2 < \rho_2)\right] = |A_1|^2(1-e^{-\rho_2}), \\ P_i &= \mathrm{E}\left[I(|w_i|^2 < \rho_{i+1}) \prod_{s=1}^{i-1} I(|w_s|^2 \geq \rho_{s+1})\right], \\ \tilde{P}_i &= \mathrm{E}\left[\prod_{s=1}^{i-1} I(|w_s|^2 \geq \rho_{s+1})\right] = 1 - \sum_{k=1}^{i-1} P_k. \end{aligned} \tag{139}$$

From (135) and (138), we have

$$\mathrm{MSE}^{(N)} \times \mathrm{SNR}^{(N)} = \left(\sum_{i=1}^{N-1} \frac{T_i}{Q_i} + \frac{\tilde{T}_N}{Q_N}\right)\left(\sum_{i=1}^{N-1} Q_i P_i + Q_N \tilde{P}_N\right). \tag{140}$$

14 Appendix: Derivation of two-step minmax criteria

The ML estimate and the MSE for the two-step process described by (54) and (55) are given by

$$\begin{aligned} \widehat{\boldsymbol{\theta}}^{(2)} &= \frac{1}{\sqrt{E_0}} \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{H}(\mathbf{u}_\mathrm{m})^H \mathbf{R}_\mathbf{n}^{-1} \left(\frac{\alpha_1 \mathbf{y}_1 + \alpha_2(\mathbf{y}_1)\mathbf{y}_2}{\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)}\right), \\ \mathrm{MSE}^{(2)} &= \frac{1}{E_0} \mathrm{E}\left[(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(2)})(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(2)})^H\right] \\ &= \frac{1}{E_0} \mathrm{E}\left[\mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{H}(\mathbf{u}_\mathrm{m})^H \mathbf{R}_\mathbf{n}^{-1} \left(\frac{\alpha_1^2 \mathbf{n}_1 \mathbf{n}_1^H + \alpha_2^2(\mathbf{y}_1)\mathbf{R}_\mathbf{n}}{\left(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1)\right)^2}\right) \mathbf{R}_\mathbf{n}^{-1} \mathbf{H}(\mathbf{u}_\mathrm{m}) \mathbf{W}_{\mathbf{u}_\mathrm{m}}\right]. \end{aligned}$$

Then,

$$\begin{aligned}
\Phi(\mathbf{u}, \mathrm{MSE}^{(2)}) &= \mathbf{u}^H \mathrm{MSE}^{(2)} \mathbf{u} \\
&= \frac{1}{E_0} \mathrm{E}\left[\frac{\alpha_1^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2} \left| \mathbf{u}^H \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{H}(\mathbf{u}_\mathrm{m})^H \mathbf{R}_\mathbf{n}^{-1} \mathbf{n}_1 \right|^2 \right. \\
&\quad \left. + \frac{\alpha_2^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2} \mathbf{u}^H \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{u} \right] \\
&= \frac{1}{E_0} \mathbf{u}^H \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{u} \; \mathrm{E}\left[\frac{\alpha_1^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2} \left| \frac{\mathbf{u}^H \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{H}(\mathbf{u}_\mathrm{m})^H \mathbf{R}_\mathbf{n}^{-1} \mathbf{n}_1}{\sqrt{\mathbf{u}^H \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{u}}} \right|^2 \right. \\
&\quad \left. + \frac{\alpha_2^2}{(\alpha_1^2 + \alpha_2^2(\mathbf{y}_1))^2} \right] \\
&= \Phi(\mathbf{u}, \mathrm{MSE}^{(1)}) \; \mathrm{E}\left[\frac{\alpha_1^2 \left| \tilde{n}_1(\mathbf{y}_1; \boldsymbol{\theta}) \right|^2 + \alpha_2^2(\tilde{n}_1(\mathbf{y}_1; \boldsymbol{\theta}))}{(\alpha_1^2 + \alpha_2^2(\tilde{n}_1(\mathbf{y}_1; \boldsymbol{\theta})))^2} \right], \qquad (141)
\end{aligned}$$

where

$$\tilde{n}_1(\mathbf{y}_1; \boldsymbol{\theta}) = \frac{\mathbf{u}^H \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{H}(\mathbf{u}_\mathrm{m})^H \mathbf{R}_\mathbf{n}^{-1} (\mathbf{y}_1 - \mathbf{H}(\mathbf{x}_1)\boldsymbol{\theta})}{\sqrt{\mathbf{u}^H \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{u}}} = \frac{\mathbf{u}^H \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{H}(\mathbf{u}_\mathrm{m})^H \mathbf{R}_\mathbf{n}^{-1} \mathbf{n}_1}{\sqrt{\mathbf{u}^H \mathbf{W}_{\mathbf{u}_\mathrm{m}} \mathbf{u}}} \qquad (142)$$

is distributed $\mathcal{CN}(0, 1)$.

15 Appendix: Distribution of Gaussian Mixture

Let $Y = \sum_{i=1}^N \alpha_i X_i^2$ where $\{X_i\}_{i=1}^N$ are independent unit normal random variables. Denote $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_N]$. Then the density ($g_N(\boldsymbol{\alpha}, y)$) and distribution ($G_N(\boldsymbol{\alpha}, y)$) function of Y is given as [46]

$$g_N(\boldsymbol{\alpha}, y) = \left(\frac{y}{2}\right)^{\frac{n}{2}-1} \sum_{k=0}^{\infty} \frac{c_k (-1)^k \left(\frac{y}{2}\right)^k}{2\Gamma(\frac{n}{2} + k)} \qquad (143)$$

$$G_N(\boldsymbol{\alpha}, y) = \left(\frac{y}{2}\right)^{\frac{n}{2}} \sum_{k=0}^{\infty} \frac{c_k (-1)^k \left(\frac{y}{2}\right)^k}{\Gamma(\frac{n}{2} + k + 1)}, \qquad (144)$$

where c_k are determined by

$$c_0 = \Pi_{j=1}^{n} \alpha_j^{-\frac{1}{2}} \tag{145}$$

$$d_k = \frac{1}{2} \sum_{j=1}^{n} \alpha_j^{-k}, \ k \geq 1 \tag{146}$$

$$c_k = \frac{1}{k} \sum_{r=0}^{k-1} d_{k-r} c_r \tag{147}$$

Since $w_1 = \frac{1}{N} \sum_{i=1}^{N} |\widehat{\mathbf{n}}_{1,i}|^2 = \frac{1}{2N} \sum_{i=1}^{2N} X_i^2$ where $\{X_i = \sqrt{2}\mathrm{Re}(\widehat{\mathbf{n}}_{1,i})\}_{i=1}^{N}$ and $\{X_i = \sqrt{2}\mathrm{Im}(\widehat{\mathbf{n}}_{1,i})\}_{i=N+1}^{2N}$ are independent unit normal random variables. $\boldsymbol{\alpha} = \frac{1}{2N}[1, \ldots, 1]$ yields

$$c_k = \binom{N+k-1}{k} (2N)^{N+k}, \ k \geq 0 \tag{148}$$

and the distribution of w_1 can then be written as

$$f_{w_1}(y) = \frac{N^N}{(N-1)!} y^{n-1} e^{-Ny}, \ y \geq 0 \tag{149}$$

$$F_{w_1}(y) = 1 - e^{-Ny} \sum_{j=0}^{N} \frac{(Ny)^j}{j!}, \tag{150}$$

Note that when all α_i's are equal, we have a mean of $2N$ independent unit normal random variables equivalent to a sum of N independent exponential random variables which is indeed a gamma distribution.

16 Appendix: Proof of Theorem 7.1

Proof. Since the shape of the design vectors $\mathbf{x}_i$ is $\mathbf{v}_{\mathrm{m}}$, we can write the set of transmitted signals as

$$\mathbf{x}_i \left(\mathbf{y}_1, \ldots, \mathbf{y}_{i-1}\right) = \mathbf{v}_{\mathrm{m}} \sqrt{E_0} \alpha_i \left(\mathbf{y}_1 - \mathbf{h}_1(\mathbf{x}_1)\theta_1, \ldots, \mathbf{y}_{i-1} - \mathbf{h}_1(\mathbf{x}_{i-1})\theta_1\right), \quad i = 1, 2, \ldots, K. \tag{151}$$

The inherent problem with any design $\mathcal{S}$ is the fact that the transmitted signal depend on the past through the noise magnitudes in the previous stages, i.e., we need to know the value of θ_1 to achieve the optimal performance. We overcome the dependence on the parameter of interest using a strategy

similar to the one in Section 12.1. We replace θ_1 with a guess of θ_1 namely θ_g in the solution in (151). Then we have,

$$\begin{aligned}\mathbf{x}_i &= \mathbf{v}_{\mathrm{m}}\sqrt{E_0}\alpha_i\left(\{\mathbf{y}_k - \mathbf{h}_1(\mathbf{x}_k)\theta_g\}_{k=1}^{i-1}\right) \\ &= \mathbf{v}_{\mathrm{m}}\sqrt{E_0}\alpha_i\left(\left\{\mathbf{n}_k + \sqrt{E_0}\mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\alpha_k(\theta_1-\theta_g)\right\}_{k=1}^{i-1}\right) \\ &= \mathbf{v}_{\mathrm{m}}\sqrt{E_0}\alpha_i\left(\{\mathbf{n}_k + \alpha_k\mathbf{z}\}_{k=1}^{i-1}\right), \end{aligned} \tag{152}$$

where $\mathbf{z} = \mathbf{h}_1(\mathbf{v}_{\mathrm{m}})\sqrt{E_0}(\theta_1 - \theta_g)$. Then the MSE of this N-step procedure can be written as

$$\mathrm{MSE}^{(K)}(\mathbf{z}) = \frac{\sigma^2}{\lambda_{\mathrm{m}}E_0}\mathrm{E}\left[\frac{\left|\sum_{i=1}^K \alpha_i\left(\{\mathbf{n}_k + \alpha_k\mathbf{z}\}_{k=1}^{i-1}\right)\widetilde{n}_i\right|^2}{\left(\sum_{i=1}^K \alpha_i^2\left(\{\mathbf{n}_k + \alpha_k\mathbf{z}\}_{k=1}^{i-1}\right)\right)^2}\right] \tag{153}$$

under the average constraint given by

$$\mathrm{SNR}^{(K)}(\mathbf{z}) = \mathrm{SNR}_0\mathrm{E}\left[\sum_{i=1}^K \alpha_i^2\left(\{\mathbf{n}_k + \alpha_k\mathbf{z}\}_{k=1}^{i-1}\right)\right] \leq \mathrm{SNR}_0. \tag{154}$$

Denote $\eta^{(K)}(\mathbf{z})$ as

$$\eta^{(K)}(\mathbf{z}) = \mathrm{MSE}^{(K)}(\mathbf{z}) \times \mathrm{SNR}^{(K)}(\mathbf{z}) \tag{155}$$

Proposition 16.1. *$MSE^{(K)}(z)$ and $SNR^{(K)}(z)$ are continuous functions of* $\mathbf{z}$

Proof. If the functions $\{\alpha_i\left(\{\mathbf{n}_k + \mathbf{z}\}_{k=1}^{i-1}\right)_{i=1}^K$ are continuous, then it follows that $\mathrm{MSE}^{(K)}(\mathbf{z})$ and $\mathrm{SNR}^{(K)}(\mathbf{z})$ are continuous since $\mathrm{E}[\cdot]$ and pdf of $\{\mathbf{n}_k\}_{k=1}^K$ are continuous functions. Also if the solution to $\{\alpha_i\}_{i=1}^K$ are thresholding functions of the form (42) then the $\mathrm{MSE}^{(K)}(\mathbf{z})$ and $\mathrm{SNR}^{(K)}(\mathbf{z})$ are integrals of the probability density function of independent gaussian random variable over ellipsoids whose center is given through $\mathbf{z}$ and hence are still continuous functions in $\mathbf{z}$. It then implies that $\eta^{(K)}(\mathbf{z})$ is also continuous in $\mathbf{z}$. □

16.0.1 $N \times K$-step procedure

Similar to the $N \times 2$-step procedure, we now construct an $N \times K$-step process, where we assume that the average energy in each of the N steps

equals E_0/N. Then the ML estimate of θ_1 for the k^{th} K-step procedure is given by

$$\widehat{\theta}_1^{(K),k} = \frac{\sum_{j=1}^{N}\{\mathbf{h}_1(\mathbf{x}_j^k\}^H\mathbf{y}_j^k}{\sum_{j=1}^{N}\|\mathbf{h}_1(\mathbf{x}_j^k)\|^2}, \tag{156}$$

where the input design vector is

$$\mathbf{x}_j^k = \mathbf{v}_{\text{m}}\sqrt{\frac{E_0}{N(1+\epsilon)}}\alpha_j\left(\left\{\mathbf{n}_i^k + \frac{\alpha_i\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right\}_{i=1}^{j-1}\right),\ 1\le i\le N,\ 1\le j\le K \tag{157}$$

and $\{\mathbf{n}_i^k\}_{i=1,k=1}^{K,N}$ are independent complex normal noises generated at the k^{th} step on the i^{th} stage. Then the overall ML estimate of θ_1 for the $N\times K$-step procedure is given by

$$\widehat{\theta}_1^{(K),N} = \frac{1}{N}\sum_{i=1}^{N}\widehat{\theta}_1^{(K),i} \tag{158}$$

and the corresponding MSE is

$$\text{MSE}^{(K),N}(\mathbf{z}) = \text{E}\left[\|\theta_1 - \widehat{\theta}_1^{(K),N}\|^2\right] = \frac{1}{N^2}\sum_{i=1}^{N}\text{E}\left[\|\theta_1 - \widehat{\theta}_1^{(K),i}\|^2\right] = \frac{1}{N}\text{MSE}^{(K),1}(\mathbf{z}), \tag{159}$$

where $\text{MSE}^{(K),1}(\mathbf{z})$ indicates the MSE of the first K^{th} estimator from the N stages and is given by

$$\begin{aligned}\text{MSE}^{(K),1}(\mathbf{z}) &= \text{E}\left[\left\|\theta_1 - \widehat{\theta}_1^{(K),1}\left(\alpha_j\left(\left\{\mathbf{n}_k^1 + \frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right\}_{k=1}^{j-1}\right)\right)\right\|^2\right] &(160)\\ &= N(1+\epsilon)\text{MSE}^{(K)}\left(\frac{\alpha_k\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right) &(161)\end{aligned}$$

Substituting the expression for $\text{MSE}^{(K),1}(\mathbf{z})$ in (161), we obtain

$$\text{MSE}^{(K),N}(\mathbf{z}) = (1+\epsilon)\text{MSE}^{(K)}\left(\frac{\mathbf{z}}{N(1+\epsilon)}\right) \tag{162}$$

The SNR of the $N \times K$-step procedure is

$$\begin{aligned}
\mathrm{SNR}^{(K),N}(\mathbf{z}) &= \sum_{j=1}^{N} \mathrm{SNR}^{(K),j} \\
&= \sum_{j=1}^{N}\sum_{i=1}^{K} \mathrm{E}\left[\left\| \mathbf{x}_i^j \left(\left\{ \mathbf{n}_k^j + \frac{\alpha_k \mathbf{z}}{\sqrt{N(1+\epsilon)}} \right\}_{k=1}^{i-1} \right) \right\|^2 \right] \\
&= \sum_{j=1}^{N} \frac{\mathrm{SNR}_0}{N(1+\epsilon)} \mathrm{E}\left[\sum_{i=1}^{K} \alpha_i^2 \left(\left\{ \mathbf{n}_k^j + \frac{\alpha_k \mathbf{z}}{\sqrt{N(1+\epsilon)}} \right\}_{k=1}^{i-1} \right) \right] \\
&= \frac{\mathrm{SNR}_0}{(1+\epsilon)} \mathrm{E}\left[\sum_{i=1}^{K} \alpha_i^2 \left(\left\{ \mathbf{n}_k^1 + \frac{\alpha_k \mathbf{z}}{\sqrt{N(1+\epsilon)}} \right\}_{k=1}^{i-1} \right) \right] \\
&= \frac{1}{(1+\epsilon)} \mathrm{SNR}^{(K)}\left(\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}} \right). \qquad (163)
\end{aligned}$$

From (162),(163), and (155), it follows that

$$\begin{aligned}
\mathrm{MSE}^{(K),N}(\mathbf{z})\mathrm{SNR}^{(K),N}(\mathbf{z}) &= \mathrm{MSE}^{(K)}\left(\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}} \right) \mathrm{SNR}^{(K)}\left(\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}} \right) \\
&= \eta^{(K)}\left(\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}} \right) \qquad (164)
\end{aligned}$$

Using continuity of $\mathrm{SNR}^{(K)}(\mathbf{z})$, for $\theta_1 \in [\theta_{\min}, \theta_{\max}]$, we have for any $\zeta_1 > 0$, $\exists \delta_1 > 0$ such that for $\left| \frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}} \right| \leq \delta_1$ we have $|\mathrm{SNR}^{(K),N}(\mathbf{z}) - \frac{\mathrm{SNR}_0}{(1+\epsilon)}| \leq \epsilon_1$. Choosing $\zeta_1 = \frac{\epsilon}{(1+\epsilon)}\mathrm{SNR}_0$, we obtain

$$\frac{1-\epsilon}{(1+\epsilon)}\mathrm{SNR}_0 \leq \mathrm{SNR}^{(K),N}(\mathbf{z}) \leq \mathrm{SNR}_0. \qquad (165)$$

The condition $\left| \frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}} \right| \leq \delta_1$ is equivalent to $N \geq N_0$, where $N_0 = \left\lceil \frac{\lambda_{\mathrm{m}} E_0 |\theta_{\max} - \theta_{\min}|^2}{\delta_1 (1+\epsilon)} \right\rceil$. Similarly continuity of $\eta^{(K)}(\mathbf{z})$ yields, for any $\zeta_2 > 0$, $\exists \delta_2 > 0$ such that for $\left| \frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}} \right| \leq \delta_2$, or equivalently $N \geq N_1$, where

$N_1 = \left\lceil \frac{\lambda_{\rm m} E_0 |\theta_{\max} - \theta_{\min}|^2}{\delta_2(1+\epsilon)} \right\rceil$, we have

$$|\eta^{(K)}\left(\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right) - \eta^*| \leq \zeta_2. \tag{166}$$

Choose $\zeta_2 = \epsilon_2 \eta^*$, then $\exists N \geq \max(N_0, N_1)$, such that

$$\begin{aligned} \mathrm{MSE}^{(K),N}(\mathbf{z}) &= \frac{\eta^{(K)}\left(\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right)}{\mathrm{SNR}^{(K),N}(\mathbf{z})} \qquad (167) \\ &\leq \frac{(1+\epsilon_2)\eta^*}{\mathrm{SNR}^{(K),N}(\mathbf{z})} \\ &\leq \frac{(1+\epsilon_2)(1+\epsilon)}{1-\epsilon}\frac{\eta^*}{\mathrm{SNR}_0} \\ &\leq (1+\delta)\frac{\eta^*}{\mathrm{SNR}_0}, \qquad (168) \end{aligned}$$

where $\delta = (1+\epsilon_2)\frac{(1+\epsilon)}{(1-\epsilon)} - 1$. Further

$$\begin{aligned} \mathrm{MSE}^{(K),N}(\mathbf{z}) &= \frac{\eta^{(K)}\left(\frac{\mathbf{z}}{\sqrt{N(1+\epsilon)}}\right)}{\mathrm{SNR}^{(K),N}(\mathbf{z})} \\ &\geq (1-\epsilon_2)\frac{\eta^*}{\mathrm{SNR}_0}. \qquad (169) \end{aligned}$$

Since $1+\delta > 1+\epsilon_2$ implies $\delta > \epsilon_2$ implies $1-\epsilon_2 > 1-\delta$. Therefore from (168) and (169),

$$\left|\mathrm{MSE}^{(K),N}(\mathbf{z}) - \frac{\eta^*}{\mathrm{SNR}_0}\right| \leq \delta \tag{170}$$

Hence we can achieve the performance of a K-step design asymptotically using an $N \times K$-step design strategy. However in practice, an $N \times K$-step procedure might be an overkill to achieve the optimal performance ϵ close. This is because for the K-step strategy, the energy at each time step is already scaled by a factor proportional to $1/K$ which implies the effect of the unknown parameter θ_1 has a reduced effect on the overall performance even without the $N \times K$-step approach. □

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, "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, "Optimal experimental design for an inverse scattering problem," *Proc. IEEE Intl. Conf. Acoust., Speech, Signal Proc.*, vol. 4, pp. 1117–1120, 2005.

[43] ——, "Sequential design for estimation and detection in linear and multiplicative models," *Technical Report, University of Michigan, Ann Arbor*, 2006.

[44] B. K. Shah and C. G. Khatri, "Distribution of a definite quadratic form for non-central normal variables," *Ann. Math. Statist.*, vol. 32, no. 3, pp. 883–887, 1961.

[45] H. Rubens, "Probability content of regions under spherical normal distributions. iv. the distribution of homogeneous and non homogeneous quadratic functions of normal variables," *Ann. Math. Statist.*, vol. 33, no. 2, pp. 542–570, 1962.

[46] S. Kotz, N. L. Johnson, and D. W. Boyd, "Series representations of distributions of quadratic forms in normal variables. i. central case," *Ann. Math. Statist.*, vol. 38, no. 3, pp. 823–837, 1967.

[47] A. C. Martinez and F. L. . Blazquez, "Distribution of a sum of weighted noncentral chi-square variables," *Sociedad de Estadistica e Investigacion Operativa Test*, vol. 14, no. 2, pp. 397–415, 2005.

[48] S. M. Kay, *Statistical Estimation.* N.J.: Prentice-Hall, Englewood-Cliffs, 1991.

[49] A. O. Hero, *Statistical Methods for Signal Processing.* Ann Arbor, MI: Course Notes, 2005.

[50] I. Telatar, "Capacity of multiple antenna gaussian channels," *Eur. Trans. Telecommun.*, vol. 10, no. 6, pp. 585–95, 1999.

[51] 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.

[52] 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.

[53] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, pp. 1451–1458, 1998.

[54] 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.

[55] C. Buianu and L. Tong, "Channel estimation for space-time orthogonal block codes," *IEEE Trans. Signal Process.*, vol. 50, pp. 2515–2528, 2002.

[56] T. L. Marzetta, "Blast training: Estimating channel characteristics for high capacity space-time wireless," *Proc. 37th Annual Allerton Conf. Communications, Control, Computing*, 1999.

[57] Y. Li, "Optimal training sequences for ofdm systems with multiple transmite antennas," *Proc. GLOBECOM*, vol. 3, pp. 1478–1482, 2000.

[58] A. Scaglione and A. Vosoughi, "Turbo estimation of channel and symbols in precoded mimo systems," *Proc. ICASSP*, vol. 4, pp. 413–416, 2004.

[59] E. G. Larsson and P. Stoica, *Space-Time Block Coding for Wireles Communications*. Cambridge, U.K: Cambridge Univ. Press, 2003.

[60] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, 1998.

[61] 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.

[62] M. Fink and C. Pradha, "Acoustic time-reversal mirrors," *Inverse Problems*, vol. 17, pp. 1–38, 2001.

[63] 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.

[64] 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.

[65] M. Fink, "Time reversal acoustics," *Physics Today*, pp. 34–40, 1997.

[66] D. Bliss, K. W. Forysthe, A. O. Hero, and A. F. Yegulalp, "Environmental issues for mimo capacity," *Signal Processing*, vol. 50, no. 9, pp. 2128–2142, 2002.