

Sensing Method for Two-Target Detection in Time-Constrained Vector Gaussian Channel

Muhammad Fahad, *Member, IEEE*
Department of Applied Computing
Michigan Technological University
Houghton, MI 49931, USA
mfahad@mtu.edu

Daniel R. Fuhrmann, *Fellow, IEEE*
Department of Applied Computing
Michigan Technological University
Houghton, MI 49931, USA
fuhrmann@mtu.edu

Abstract

This paper considers a vector Gaussian channel of fixed identity covariance matrix and binary input signalling as the mean of it. A linear transformation is performed on the vector input signal. The objective is to find the optimal scaling matrix, under the total time constraint, that would: i) maximize the mutual information between the input and output random vectors, ii) maximize the MAP detection. It was found that the two metrics lead to different optimal solutions for our experimental design problem. We have used the Monte Carlo method for our computational work.

Index Terms

sensor scheduling, vector Gaussian channels.

1. INTRODUCTION

In [1] a Gaussian channel $Y|X \sim \mathcal{N}(\sqrt{T} \cdot X, 1)$ is considered and discovered that mutual information $I(T)$ is concave function in T for arbitrary input distribution. Whereas in [2] the Poisson channel $Y|X \sim \text{Poiss}(T \cdot X)$ is investigated and found a result similar to the Gaussian channel: $I(T)$ is concave in T for arbitrary input distribution. In [3][4] it was observed that concavity of I looked preserved under linear time constraint in a vector Poisson channel with a 2-long binary input signalling and a 3-long conditionally Poisson vector. It was further observed from a computational viewpoint that MAP detector was not necessarily reaching to the same optimal argument as that was given by mutual information. Here we construct an analogous model to that of Poisson channel such that at least the concavity of I remains intact for the Gaussian channel too. Compared to vector Poisson channel, the literature on vector Gaussian channel is comparatively richer and may help in providing some insight into the Poisson channel.

In past work [5] a generalization of Bregman divergence is developed to unify the vector Gaussian and Poisson channel models from the perspective of the gradient of mutual information; and mutual information is considered for signal recovery and classification with an energy constraint $\text{Tr}(\Phi\Phi^\top) = 1$. MAP estimation is used for the classification purpose in [5] using Monte Carlo method to first approximate the gradient and then gradient descent is employed for the classification problem. It was noted in [5] that mutual information well served the classification problem. In this paper we attempted to use the detection theory criterion for signal classification and then compared with the information theoretic solution. Another work [6] provides some results relevant to Gaussian channels about the concavity of I w.r.t squared singular values of the scaling matrix when certain conditions on the channel covariance and precoder matrix are satisfied. It is found that for

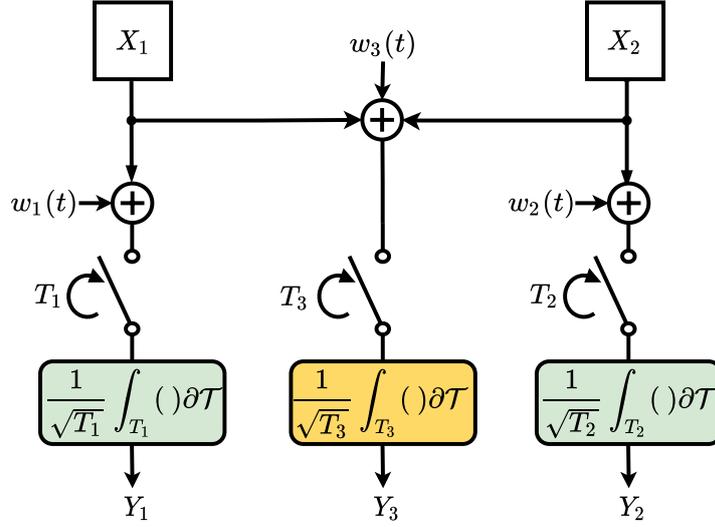


Fig. 1: Illustration of sensing paradigm for detection of 2-long hidden random vector X from 3-long observable random vector Y through a vector Gaussian channel under a total time constraint $T = \sum_{i=1}^3 T_i$. $w_i(t)$ are independent white Gaussian noise processes. Only one of the integrators becomes active for a time T_i such that time constraint is always satisfied after the total sensing time T is consumed. Objective is to maximize the mutual information between input and output, $I(X_1, X_2; Y_1, Y_2, Y_3)$, and Bayes probability of total detections, P_d , by satisfying the time constraint.

our problem I is concave in affine space defined by $(T_1, T_2, T_3) := (\frac{T-T_3}{2}, \frac{T-T_3}{2}, T_3)$ where $0 \leq T_3 \leq T$. For a Gaussian channel $Y|X \sim \mathcal{N}(\Phi X, \sigma_n^2 I)$ with input $X \sim \mathcal{N}(0, \Sigma)$ where Σ is full rank covariance matrix; then the two solutions from maximizing the mutual information and minimizing the mean-square error in scaling matrix under the power constraint $\text{Tr}(\Phi \Phi^\top)$ leads to the same optimal solution which is a water-filling power allocation i-e concentrate more power resource to modes that provide higher snr [7]. Our problem is different in the input signalling, and we took the detection theory criterion instead of the estimation theory (MMSE) and then compared the optimal solution with one obtained from I using Monte Carlo computational method.

The rest of the paper is organized as follows: Section 2 introduces the problem description, explaining the vector Gaussian channel under consideration. Section 3 provides the *information theoretic* model, while Section 4 describes the *detection theoretic* model of the problem. Section 5 discusses the computed results from the previous two sections. Finally, Section 6 concludes the paper.

Notation: Upper case letters denote random vectors. Realizations of the random vectors are denoted by lower case letters. A number in subscript is used to show the component number of the random vector. We use X_1 and Y_1 to represent scalar input and output random variables, respectively. The superscript $(\cdot)^\top$ denotes the matrix/vector transpose. T is a given finite time. α is an arbitrary positive scalar variable. Φ represents the scaling matrix. p is the prior probability and $f_X(x)$ denotes the probability mass function of X . K -dimensional multivariate Gaussian distribution is represented by $\mathcal{N}(x; \mu, \Sigma) = (2\pi)^{-\frac{k}{2}} \det(\Sigma)^{-\frac{1}{2}} e^{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)}$.

2. PROBLEM DESCRIPTION

We consider the vector Gaussian channel:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \end{bmatrix} = \begin{bmatrix} \sqrt{T_1} & 0 \\ 0 & \sqrt{T_2} \\ \sqrt{T_3} & \sqrt{T_3} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} + \begin{bmatrix} N_1 \\ N_2 \\ N_3 \end{bmatrix}, \quad (1)$$

with X_1 and X_2 be two independent and identical distributed (i.i.d) transformed Bernoulli random variables with p being the probability of occurrence of 1. We consider probability mass function f of discrete random vector $X \equiv [X_1, X_2]^T$ as

$$f_X(x) = \begin{cases} p^2 & x = [\lambda_1 \ \lambda_1]^T \\ (1-p)^2 & x = [\lambda_0 \ \lambda_0]^T \\ p(1-p) & x = [\lambda_0 \ \lambda_1]^T \text{ or } x = [\lambda_1 \ \lambda_0]^T \end{cases} \quad (2)$$

Noise vector N is a multivariate Gaussian with zero mean and identity covariance matrix; and independent of input X . The constraint on the scaling matrix is $T_1 + T_2 + T_3 = T$. The conditional distribution of vector Y given X is a multivariate Gaussian:

$$Y \Big| (X = x_1 x_2) \sim \mathcal{N} \left(\begin{bmatrix} \sqrt{T_1} \cdot x_1 \\ \sqrt{T_2} \cdot x_2 \\ \sqrt{T_3} \cdot (x_1 + x_2) \end{bmatrix}, I \right). \quad (3)$$

The objective is optimal time-allocation, (T_1, T_2, T_3) , of total available time resource, T , that would maximize the *reward* i.e. either the mutual information or probability of total correct detections. Mathematically we may write

$$\max_{T_1, T_2, T_3} I(X_1, X_2; Y_1, Y_2, Y_3) \text{ s.t. } T_1 + T_2 + T_3 = T. \quad (4)$$

From the *detection theoretic* aspect we maximize the Bayesian probability of total correct detections, P_d , of hidden random vector X from observable random vector Y , as

$$\max_{T_1, T_2, T_3} P_d \text{ s.t. } T_1 + T_2 + T_3 = T \quad (5)$$

3. INFORMATION THEORETIC DESCRIPTION

A. Scalar Gaussian channel

The scalar version of the Gaussian channel is first presented, and then we extend it to the vector version. We start with mutual information between a scalar random variable X_1 which is a transformed Bernoulli random variable and Y_1 is a univariate Gaussian mixture. The probability mass function of Y_1 is then given as

$$f(Y_1) = (1-p) \cdot \mathcal{N}(Y_1; \lambda_0 \sqrt{T}, 1) + p \cdot \mathcal{N}(Y_1; \lambda_1 \sqrt{T}, 1).$$

The mutual information I can be written as

$$I(X_1; Y_1) = H(Y_1) - H(Y_1|X_1)$$

where $H(\cdot)$ is the Shannon entropy and f is the probability mass function of random variate Y with \mathcal{Y} as the corresponding support. We may write differential entropy $H(Y_1)$ as

$$H(Y_1) = - \int_{-\infty}^{\infty} \left((1-p) \cdot \mathcal{N}(y_1; \lambda_0 \sqrt{T}, 1) + p \cdot \mathcal{N}(y_1; \lambda_1 \sqrt{T}, 1) \cdot \log_2 \left[(1-p) \cdot \mathcal{N}(y_1; \lambda_0 \sqrt{T}, 1) + p \cdot \mathcal{N}(y_1; \lambda_1 \sqrt{T}, 1) \right] \right) dy_1, \quad (6)$$

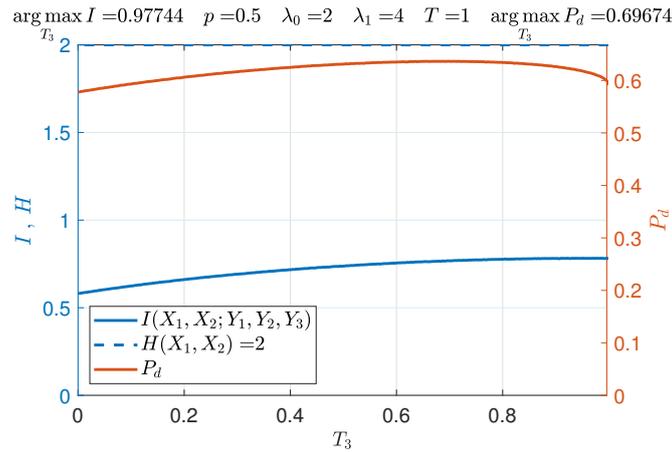


Fig. 2: Mutual information $I(X; Y)$ vs. T_3 and probability of total correct detections P_d vs. time T_3 in a time constraint $T_1 + T_2 + T_3 = 1$ such that $(T_1, T_2, T_3) := (\frac{1-T_3}{2}, \frac{1-T_3}{2}, T_3)$ where $0 \leq T_3 \leq 1$.

and

$$H(Y_1|X_1) = (1 - p) \cdot 0.5 \cdot \text{Log}_2[2\pi e] + p \cdot 0.5 \cdot \text{Log}_2[2\pi e]. \quad (7)$$

In the following section we formulate the mutual information expression for our vector Gaussian model.

B. Vector Gaussian channel

Mutual information between two random vectors can be defined as the difference between the total differential entropy in one random vector and the conditional differential entropy in the second random vector given the first vector. We write

$$I(X; Y) = H(Y) - H(Y|X) \quad (8)$$

The conditional entropy $H(Y|X)$ is calculated from the conditional probability mass functions $f(Y|X_i)$ defined as

$$\begin{aligned} f(Y|X = [\lambda_0 \quad \lambda_0]^T) &= \mathcal{N}(Y_1; \lambda_0\sqrt{T_1}, 1) \cdot \mathcal{N}(Y_2; \lambda_0\sqrt{T_2}, 1) \cdot \mathcal{N}(Y_3; 2\lambda_0\sqrt{T_3}, 1), \\ f(Y|X = [\lambda_0 \quad \lambda_1]^T) &= \mathcal{N}(Y_1; \lambda_0\sqrt{T_1}, 1) \cdot \mathcal{N}(Y_2; \lambda_1\sqrt{T_2}, 1) \cdot \mathcal{N}(Y_3; (\lambda_0 + \lambda_1)\sqrt{T_3}, 1), \\ f(Y|X = [\lambda_1 \quad \lambda_0]^T) &= \mathcal{N}(Y_1; \lambda_1\sqrt{T_1}, 1) \cdot \mathcal{N}(Y_2; \lambda_0\sqrt{T_2}, 1) \cdot \mathcal{N}(Y_3; (\lambda_1 + \lambda_0)\sqrt{T_3}, 1), \\ f(Y|X = [\lambda_1 \quad \lambda_1]^T) &= \mathcal{N}(Y_1; \lambda_1\sqrt{T_1}, 1) \cdot \mathcal{N}(Y_2; \lambda_1\sqrt{T_2}, 1) \cdot \mathcal{N}(Y_3; 2\lambda_1\sqrt{T_3}, 1). \end{aligned}$$

The marginal probability mass function of Y is then given as

$$\begin{aligned} f(Y) &= (1 - p)^2 \cdot f(Y|X = [\lambda_0 \quad \lambda_0]^T) + p(1 - p) \cdot \\ & f(Y|X = [\lambda_0 \quad \lambda_1]^T) + p(1 - p) \cdot f(Y|X = [\lambda_1 \quad \lambda_0]^T) + \\ & p^2 \cdot f(Y|X = [\lambda_1 \quad \lambda_1]^T). \end{aligned} \quad (9)$$

As illustrated in fig.(1) the sensing paradigm for our problem; mutual information $I(X; Y)$ is then defined as

$$I(X; Y) = H(Y) - H(Y|X), \quad (10)$$

where $H(Y)$ is a differential entropy of our finite Gaussian mixture model (gmm) and given as

$$\begin{aligned}
 H(Y) = - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} & \left[\left((1-p)^2 \cdot \mathcal{N}(y_1; \lambda_0 T_1, 1) \cdot \mathcal{N}(y_2; \lambda_0 T_2, 1) \cdot \mathcal{N}(y_3; 2\lambda_0 T_3, 1) + \right. \right. \\
 & p(1-p) \cdot \mathcal{N}(y_1; \lambda_0 T_1, 1) \cdot \mathcal{N}(y_2; \lambda_1 T_2, 1) \cdot \mathcal{N}(y_3; (\lambda_0 + \lambda_1) T_3, 1) + \\
 & p(1-p) \cdot \mathcal{N}(y_1; \lambda_1 T_1, 1) \cdot \mathcal{N}(y_2; \lambda_0 T_2, 1) \cdot \mathcal{N}(y_3; (\lambda_1 + \lambda_0) T_3, 1) + \\
 & \left. \left. p^2 \cdot \mathcal{N}(y_1; \lambda_1 T_1, 1) \cdot \mathcal{N}(y_2; \lambda_1 T_2, 1) \cdot \mathcal{N}(y_3; 2\lambda_1 T_3, 1) \right) \cdot \right. \\
 & \left(\text{Log}_2[(1-p)^2 \cdot \mathcal{N}(y_1; \lambda_0 T_1, 1) \cdot \mathcal{N}(y_2; \lambda_0 T_2, 1) \cdot \mathcal{N}(y_3; 2\lambda_0 T_3, 1) + \right. \\
 & p(1-p) \cdot \mathcal{N}(y_1; \lambda_0 T_1, 1) \cdot \mathcal{N}(y_2; \lambda_1 T_2, 1) \cdot \mathcal{N}(y_3; (\lambda_0 + \lambda_1) T_3, 1) + \\
 & p(1-p) \cdot \mathcal{N}(y_1; \lambda_1 T_1, 1) \cdot \mathcal{N}(y_2; \lambda_0 T_2, 1) \cdot \mathcal{N}(y_3; (\lambda_1 + \lambda_0) T_3, 1) + \\
 & \left. \left. p^2 \cdot \mathcal{N}(y_1; \lambda_1 T_1, 1) \cdot \mathcal{N}(y_2; \lambda_1 T_2, 1) \cdot \mathcal{N}(y_3; 2\lambda_1 T_3, 1)] \right) \right] dy_1 dy_2 dy_3, \quad (11)
 \end{aligned}$$

and $H(Y|X)$ is

$$\begin{aligned}
 H(Y|X) = (1-p)^2 \cdot (0.5 \cdot \text{Log}_2[2\pi e] + 0.5 \cdot \text{Log}_2[2\pi e] + 0.5 \cdot \text{Log}_2[2\pi e]) + \\
 p(1-p) \cdot (0.5 \cdot \text{Log}_2[2\pi e] + 0.5 \cdot \text{Log}_2[2\pi e] + 0.5 \cdot \text{Log}_2[2\pi e]) + \\
 p(1-p) \cdot (0.5 \cdot \text{Log}_2[2\pi e] + 0.5 \cdot \text{Log}_2[2\pi e] + 0.5 \cdot \text{Log}_2[2\pi e]) + \\
 p^2 \cdot (0.5 \cdot \text{Log}_2[2\pi e] + 0.5 \cdot \text{Log}_2[2\pi e] + 0.5 \cdot \text{Log}_2[2\pi e]). \quad (12)
 \end{aligned}$$

Since multidimensional integral defined in (11) have no closed-form solution, we have to resort to numerical methods. We may mitigate the curse of dimensionality involved in multi-dimensional integration by Monte-Carlo technique by taking samples from the multivariate Gaussian mixture distribution to achieve fast convergence to the true mixture differential entropy at a reasonable computational burden; whereas naive uniform sampling of the space would lead to a quite slow convergence to the true differential entropy.

$$\begin{aligned}
 H(Y) &= E[-\text{Log}_2[f_Y(Y)]] \\
 &= - \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_Y(y) \cdot \text{Log}_2[f_Y(y)] dy \\
 &\approx - \frac{\sum_i \text{Log}_2[f_Y(s_i)]}{N_s}, \quad (13)
 \end{aligned}$$

where $f_Y(\cdot)$ is the mixture probability distribution of Y , N_s is the number of MC samples and s_i is the i^{th} sample from multivariate Gaussian mixture distribution [8].

Theorem 1: $I(X_1, X_2; Y_1, Y_2, Y_3)$ is symmetric in variables T_1 and T_2 .

Proof: Mutual information $I(X_1, X_2; Y_1, Y_2, Y_3)$ given in (10) is invariant under any permutation of variables T_1 and T_2 . That means interchanging the two variables leaves the expression unchanged. ■

Theorem 2: $I(X_1, X_2; Y_1, Y_2, Y_3)$ is concave in $T_3 = 0$ plane.

Proof:

$$I(X_1, X_2; Y_1, Y_2, Y_3) \Big|_{T_3=0} = I(X_1, X_2; Y_1, Y_2)$$

By chain rule of mutual information:

$$\begin{aligned} I(X_1, X_2; Y_1, Y_2) &= I(X_1, X_2; Y_1) + I(X_1, X_2; Y_2|Y_1) \\ &= I(X_1; Y_1) + I(X_2; Y_2). \end{aligned} \quad (14)$$

We note in (14) that each $I(X_i; Y_i)$ is solely a function of T_i and also concave in it [1]. Since the sum of concave functions is a concave function. Therefore $I(X_1, X_2; Y_1, Y_2)$ is concave in T_1 and T_2 when $T_3 = 0$. This concludes the proof. ■

Theorem 3: $I(X_1, X_2; Y_1, Y_2, Y_3)$ is concave in T_3 along the line $(T_1, T_2, T_3) := (\frac{T-T_3}{2}, \frac{T-T_3}{2}, T_3)$ parametrized by $0 \leq T_3 \leq T$.

Proof: It is noted in [9, Theorem 5] that mutual information is a concave function of the squared singular values (λ) of the precoder matrix P if the first m' eigenvectors of the channel covariance matrix ($\mathbf{R}_H = H^T \mathbf{R}_Z^{-1} H$) coincide with the left singular vectors of the precoder P i.e $H_\lambda I(S; Y) \leq 0$ for the signal model $Y = HPS + Z$ where $H \in \mathbb{R}^{n \times p}$ is the channel, S is the input signaling $S \in \mathbb{R}^m$, P is a precoder matrix $P \in \mathbb{R}^{p \times m}$ and $Z \in \mathbb{R}^n$ is Gaussian noise independent of the input S and has covariance matrix \mathbf{R}_Z .

For our problem: $H = I$, $\mathbf{R}_Z^{-1} = \Lambda = I$, $P = \Phi$, $S = X$ and $Z = N$. The singular value decomposition of $\Phi = U\Sigma V^*$. We have singular matrix

$$\Sigma = \begin{bmatrix} \frac{\sqrt{T_1+T_2+2T_3-\sqrt{(T_1-T_2)^2+4T_3^2}}}{\sqrt{2}} & 0 \\ 0 & \frac{\sqrt{T_1+T_2+2T_3+\sqrt{(T_1-T_2)^2+4T_3^2}}}{\sqrt{2}} \\ 0 & 0 \end{bmatrix} \quad (15)$$

By substituting $T_1 = T_2 = \frac{T-T_3}{2}$, the squared singular values are $[\lambda_1 \ \lambda_2 \ \lambda_3] = [\frac{T-T_3}{2} \ \frac{T+3T_3}{2} \ 0]$ for $0 \leq T_3 \leq T$. This is just the composition with an affine transformation on the domain. Concavity remains preserved under affine transformation [10, page 79-86]. ■

The above concavity of I is illustrated in fig.(2).

4. DETECTION THEORETIC DESCRIPTION

A. Bayes risk

In last section, we presented the metric of mutual information I between hidden random vector X , and observable vector Y . Here we approach the sensing problem as a multi-hypothesis detection problem and define the Bayesian risk [11, pp.220] to minimize in (T_1, T_2, T_3) . We define Bayes risk r as

$$\begin{aligned} r &= (1-p)^2 \left[P_{\lambda_0 \lambda_0 | \lambda_0 \lambda_0} C_{\lambda_0 \lambda_0 | \lambda_0 \lambda_0} + P_{\lambda_0 \lambda_1 | \lambda_0 \lambda_0} C_{\lambda_0 \lambda_1 | \lambda_0 \lambda_0} \right. \\ &\quad \left. + P_{\lambda_1 \lambda_0 | \lambda_0 \lambda_0} C_{\lambda_1 \lambda_0 | \lambda_0 \lambda_0} + P_{\lambda_1 \lambda_1 | \lambda_0 \lambda_0} C_{\lambda_1 \lambda_1 | \lambda_0 \lambda_0} \right] + p(1-p) \\ &\quad \left[P_{\lambda_0 \lambda_0 | \lambda_0 \lambda_1} C_{\lambda_0 \lambda_0 | \lambda_0 \lambda_1} + P_{\lambda_0 \lambda_1 | \lambda_0 \lambda_1} C_{\lambda_0 \lambda_1 | \lambda_0 \lambda_1} + P_{\lambda_1 \lambda_0 | \lambda_0 \lambda_1} C_{\lambda_1 \lambda_0 | \lambda_0 \lambda_1} \right. \\ &\quad \left. + P_{\lambda_1 \lambda_1 | \lambda_0 \lambda_1} C_{\lambda_1 \lambda_1 | \lambda_0 \lambda_1} \right] + p(1-p) \left[P_{\lambda_0 \lambda_0 | \lambda_1 \lambda_0} C_{\lambda_0 \lambda_0 | \lambda_1 \lambda_0} \right. \\ &\quad \left. + P_{\lambda_0 \lambda_1 | \lambda_1 \lambda_0} C_{\lambda_0 \lambda_1 | \lambda_1 \lambda_0} + P_{\lambda_1 \lambda_0 | \lambda_1 \lambda_0} C_{\lambda_1 \lambda_0 | \lambda_1 \lambda_0} \right] \\ &\quad + p^2 \left[P_{\lambda_0 \lambda_0 | \lambda_1 \lambda_1} C_{\lambda_0 \lambda_0 | \lambda_1 \lambda_1} + P_{\lambda_0 \lambda_1 | \lambda_1 \lambda_1} C_{\lambda_0 \lambda_1 | \lambda_1 \lambda_1} + P_{\lambda_1 \lambda_0 | \lambda_1 \lambda_1} C_{\lambda_1 \lambda_0 | \lambda_1 \lambda_1} \right. \\ &\quad \left. + P_{\lambda_1 \lambda_1 | \lambda_1 \lambda_1} C_{\lambda_1 \lambda_1 | \lambda_1 \lambda_1} \right], \end{aligned}$$

where $P_{\lambda_k \lambda_l | \lambda_i \lambda_j}$ is the probability that $X = [\lambda_i, \lambda_j]^T$ is true while decision $X = [\lambda_k, \lambda_l]^T$ is made; similarly for $C_{\lambda_k \lambda_l | \lambda_i \lambda_j}$. Setting all costs for which $[\lambda_i, \lambda_j]^T \neq [\lambda_k, \lambda_l]^T$ to one and $[\lambda_i, \lambda_j]^T = [\lambda_k, \lambda_l]^T$ to zero, we have

$$\begin{aligned}
 r = (1-p)^2 & \left[P_{\lambda_0 \lambda_1 | \lambda_0 \lambda_0} + P_{\lambda_1 \lambda_0 | \lambda_0 \lambda_0} + P_{\lambda_1 \lambda_1 | \lambda_0 \lambda_0} \right] + p(1-p) \\
 & \left[P_{\lambda_0 \lambda_0 | \lambda_0 \lambda_1} + P_{\lambda_1 \lambda_0 | \lambda_0 \lambda_1} + P_{\lambda_1 \lambda_1 | \lambda_0 \lambda_1} \right] + p(1-p) \\
 & \left[P_{\lambda_0 \lambda_0 | \lambda_1 \lambda_0} + P_{\lambda_0 \lambda_1 | \lambda_1 \lambda_0} + P_{\lambda_1 \lambda_1 | \lambda_1 \lambda_0} \right] + p^2 \\
 & \left[P_{\lambda_0 \lambda_0 | \lambda_1 \lambda_1} + P_{\lambda_0 \lambda_1 | \lambda_1 \lambda_1} + P_{\lambda_1 \lambda_0 | \lambda_1 \lambda_1} \right]. \tag{16}
 \end{aligned}$$

We are interested in minimizing this Bayes risk r in (T_1, T_2, T_3) i-e

$$\min_{T_1, T_2, T_3} r \quad \text{s.t. } T_1 + T_2 + T_3 = T. \tag{17}$$

Note that while minimizing r in (T_1, T_2, T_3) , the decisions boundaries would be changing accordingly and become function of (T_1, T_2, T_3) . Equivalently, we may say that

$$\max_{T_1, T_2, T_3} P_d \quad \text{s.t. } T_1 + T_2 + T_3 = T \tag{18}$$

where P_d is probability of total correct detections, $P_d = 1 - r$. In the next section we present the computed results of (18).

5. MONTE CARLO SIMULATION RESULTS

For all simulation purposes, we have assumed that optimizing argument in $\max_{T_1, T_2, T_3} I(X_1, X_2; Y_1, Y_2, Y_3)$ s.t. $T_1 + T_2 + T_3 = T$ would have $T_1 = T_2$. This is based on the observations noted in the ternary diagrams given in fig.(3). We computed I for a wide range of given input parameters λ_0, λ_1, T and p ; and it was noted that maximizing argument always seems to lie on the line $(T_1, T_2, T_3) := (\frac{T-\alpha}{2}, \frac{T-\alpha}{2}, \alpha)$ where $0 \leq \alpha \leq T$. In other words we noted a Schur concavity of I in (T_1, T_2) whenever T_3 is held fixed under a given time-constraint; however no proof of Schur concavity of I is claimed in this work.

We compute mutual information in (11) by first computing the entropy of the multivariate Gaussian mixture by generating the samples from it. Each of the Gaussian mixture component is a 3-dimensional multivariate Gaussian distribution that comes with a prior belief. We generate a total of 10^6 samples to calculate $H(Y)$ for a given prior p and energy constraint $T_1 + T_2 + T_3 = T$ with $T_1 = T_2$. Since we do have a closed-form available for a differential entropy of a multivariate Gaussian distribution therefore for $H(Y|X)$ we don't need to apply the Monte Carlo method for evaluating it. The difference of the two would provide the approximated value of $I(X; Y) \Big|_{(\frac{T-\alpha}{2}, \frac{T-\alpha}{2}, \alpha)}$ for a given set of parameters.

For the MAP detection we use the empirical method to calculate the probability of total correct detections P_d . We again assume that the optimal solution has $T_1 = T_2$. An optimal solution for any given set of parameters is then searched in the region $(T_1, T_2, T_3) := (\frac{T-\alpha}{2}, \frac{T-\alpha}{2}, \alpha)$ where $0 \leq \alpha \leq T$. For a given value of T : α takes 400 linear steps from 0 to T and for each step we first generate the samples from the Gaussian mixture under consideration by additionally knowing which mixture component has actually generated any particular sample. For every input sample we then computed the posterior probability for each of four hypotheses and then decide in favor of the hypothesis that has the highest posterior probability. Comparing our 10^6 decisions with that of the 10^6 inputs, we can then calculate the total correct detections for each of the

discrete α steps. This way for any given set of parameters λ_0 , λ_1 , T , p and α we may empirically compute the $P_d \Big|_{(\frac{T-\alpha}{2}, \frac{T-\alpha}{2}, \alpha)}$.

For the sake of simplicity we call the time proportion: $(\frac{T}{2}, \frac{T}{2}, 0)$ to be the individual sensing; $(0, 0, T)$ to be the joint sensing and $(\frac{T-\alpha}{2}, \frac{T-\alpha}{2}, \alpha)$ where $0 < \alpha < T$ to be hybrid sensing method. In fig.(4) and fig.(5), there are a couple of observations to be noted: first we can see the concavity of I and P_d in T_3 ; second observation is maximizing the mutual information and probability of total correct detections doesn't lead to the same optimal solution; this is more noticeable in fig.(5) where mutual information is maximizing in the vicinity of $T_3 = 1$ and therefore suggesting individual sensing to be optimum whereas probability of total correct detections is suggesting the hybrid sensing to be optimum. The third observation is that just by looking at the prior p we can't say in the most rough sense that which of the three sensing mechanisms would be optimal, either from the perspective of the mutual information or from the Bayes inference. This is unlike to that of a Poisson problem in [3] where individual sensing was always optimal whenever $p \geq 0.5$ irrespective of the given input set of parameters from mutual information perspective.

To further expand our understanding if hybrid sensing remains optimal for a wide range of input parameters λ_0 and λ_1 for fixed prior p and time constraint $T = 1$, we simulate another Monte Carlo simulation. The input parameter set is $\{(\lambda_0, \lambda_1) \in \mathbb{R}_+ \times \mathbb{R}_+ \mid 0 < \lambda_1 \leq 5 \text{ and } \lambda_1 > \lambda_0\}$. For each of (λ_0, λ_1) we compute 400 values of mutual information by varying α linearly from 0 to $T = 1$ in 400 steps. For each step 10^5 samples are used for calculation of differential entropy $H(Y)$. Scatter plots on the left-hand side in fig.(6) illustrates the respective optimal value of $I(X; Y)$ at each input parameter for the prior taking values: 0.125, 0.5 and 0.99. Whereas the scatter plots on the right-hand side illustrates the corresponding optimizing argument $(\frac{T-T_3}{2}, \frac{T-T_3}{2}, T_3)$ where $0 < T_3 < T$. It can be seen that when the two input parameters λ_0 and λ_1 are closer (as in the diagonal) the mutual information is near to zero and hybrid sensing is the best sensing strategy; as the two input parameters gets farther (as in the lower right corner in scatter plot) the mutual information gets higher and still the hybrid sensing is optimal. This is true for all three values of the prior. When the same simulation is run for maximizing the Bayes probability of total correct detections P_d the results are shown in scatter plots of fig.(7). P_d is shown on the left scatter plot for each prior. As the input parameters λ_0 and λ_1 gets closer (as in the diagonal) the P_d touches the maximum value among the $\{(1-p)^2, 2p(1-p), p^2\}$. In the lower right corner the P_d is highest as the input parameters are the farthest apart. The right scatter plots illustrates that hybrid sensing is the optimum method from the Bayes detection point of view. It must be noted that even the hybrid sensing is optimal from perspectives of the mutual information and Bayes probability of total correct detection; the optimal arguments from these two metrics are not necessarily appear to be the same. These simulations therefore constitute a counter-example where information theory and detection theory are leading to different optimal solutions.

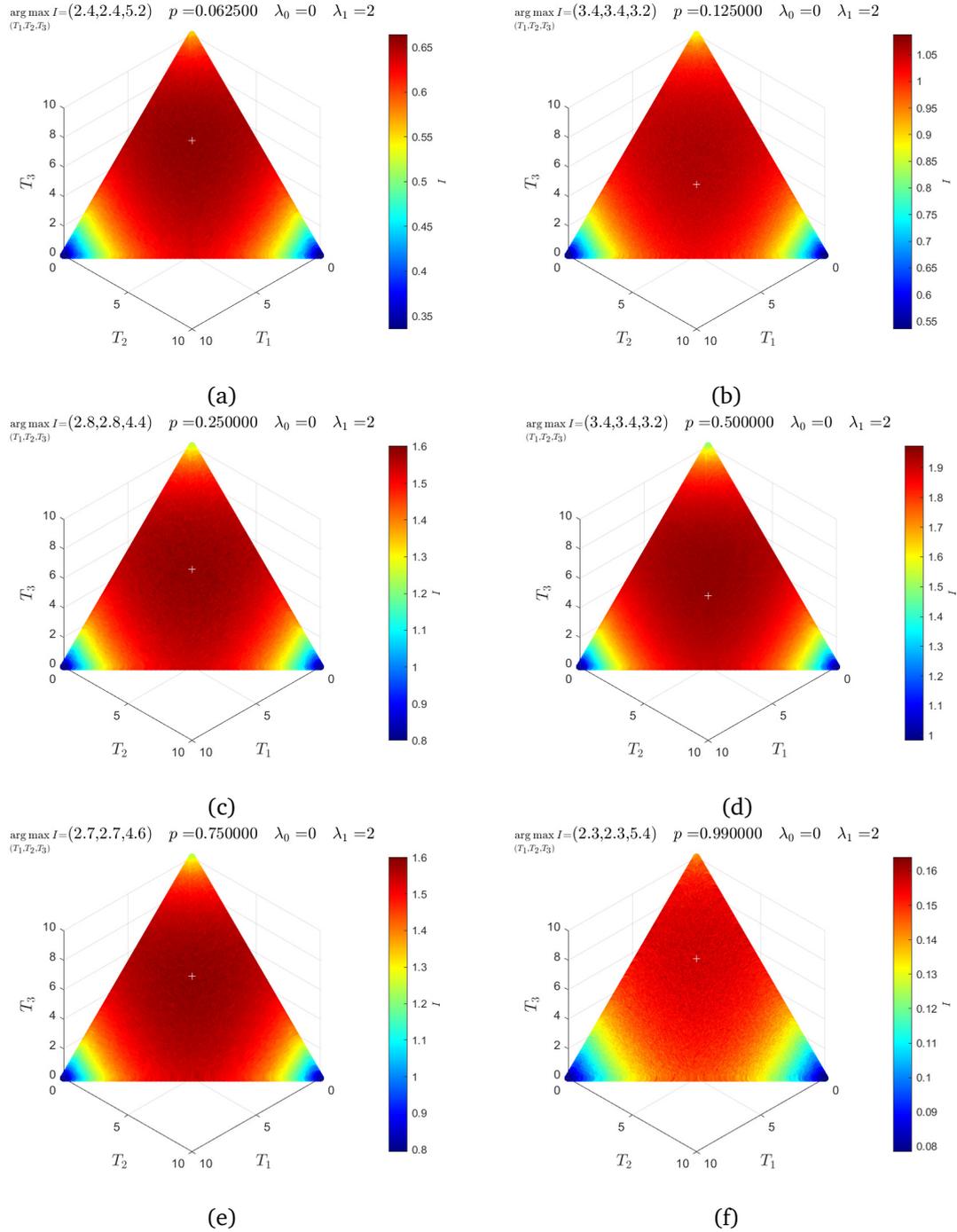


Fig. 3: $I(X; Y)$ vs. (T_1, T_2, T_3) under time constraint $T_1 + T_2 + T_3 = 10$ for $\lambda_0 = 0$, $\lambda_1 = 2$, and varying prior probability p .

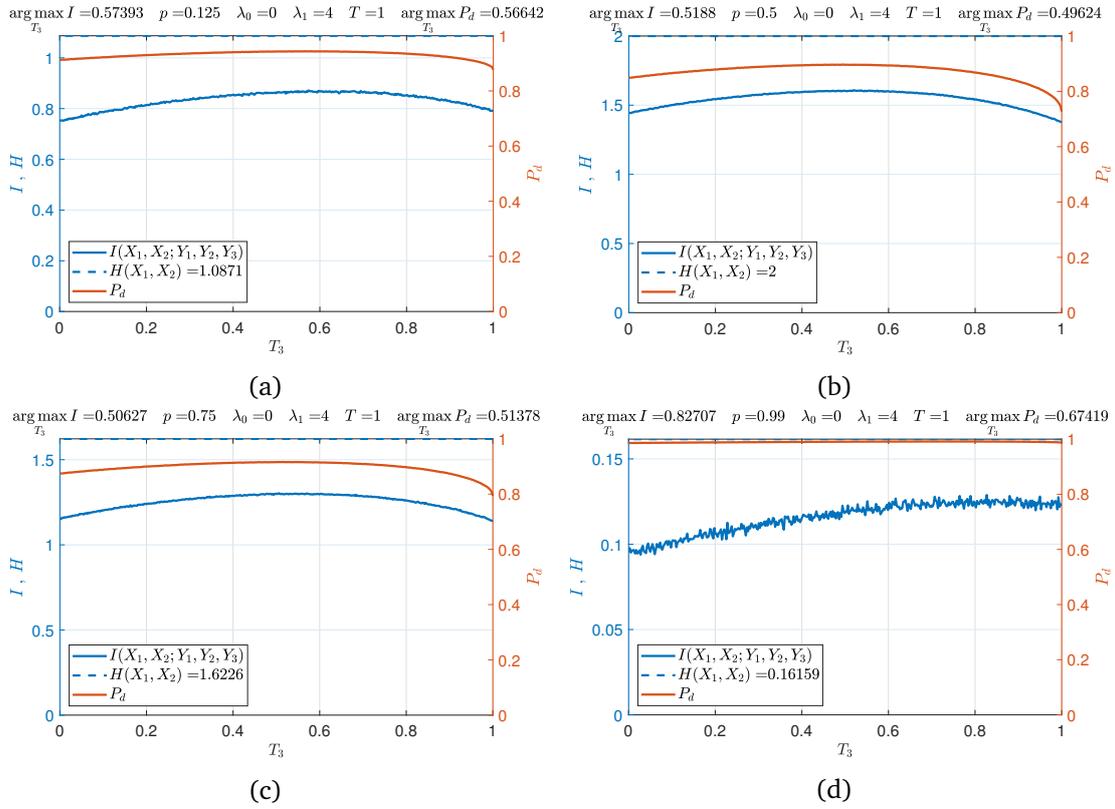


Fig. 4: Mutual information $I(X; Y)$ vs. T_3 and probability of total correct detections P_d vs. T_3 for prior probabilities of 0.125, 0.5, 0.75 and 0.99.

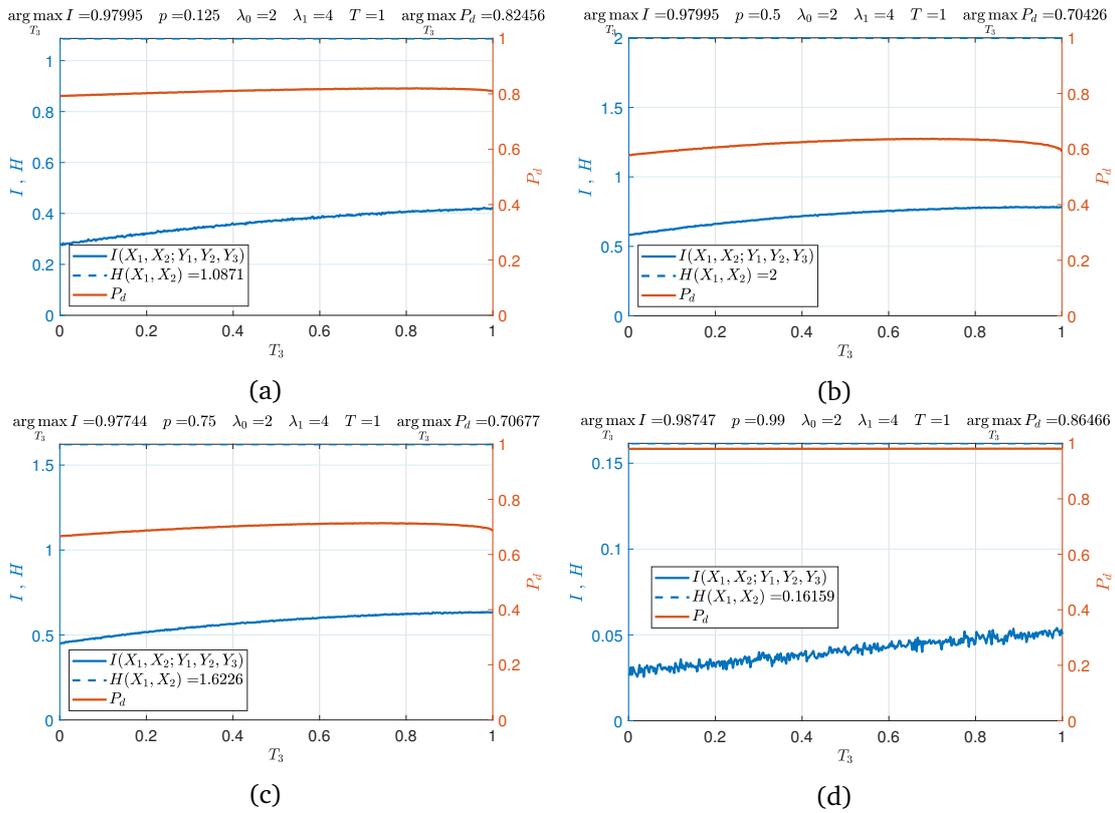


Fig. 5: Mutual information $I(X; Y)$ vs. T_3 and probability of total correct detections P_d vs. T_3 for prior probabilities of 0.125, 0.5, 0.75 and 0.99.

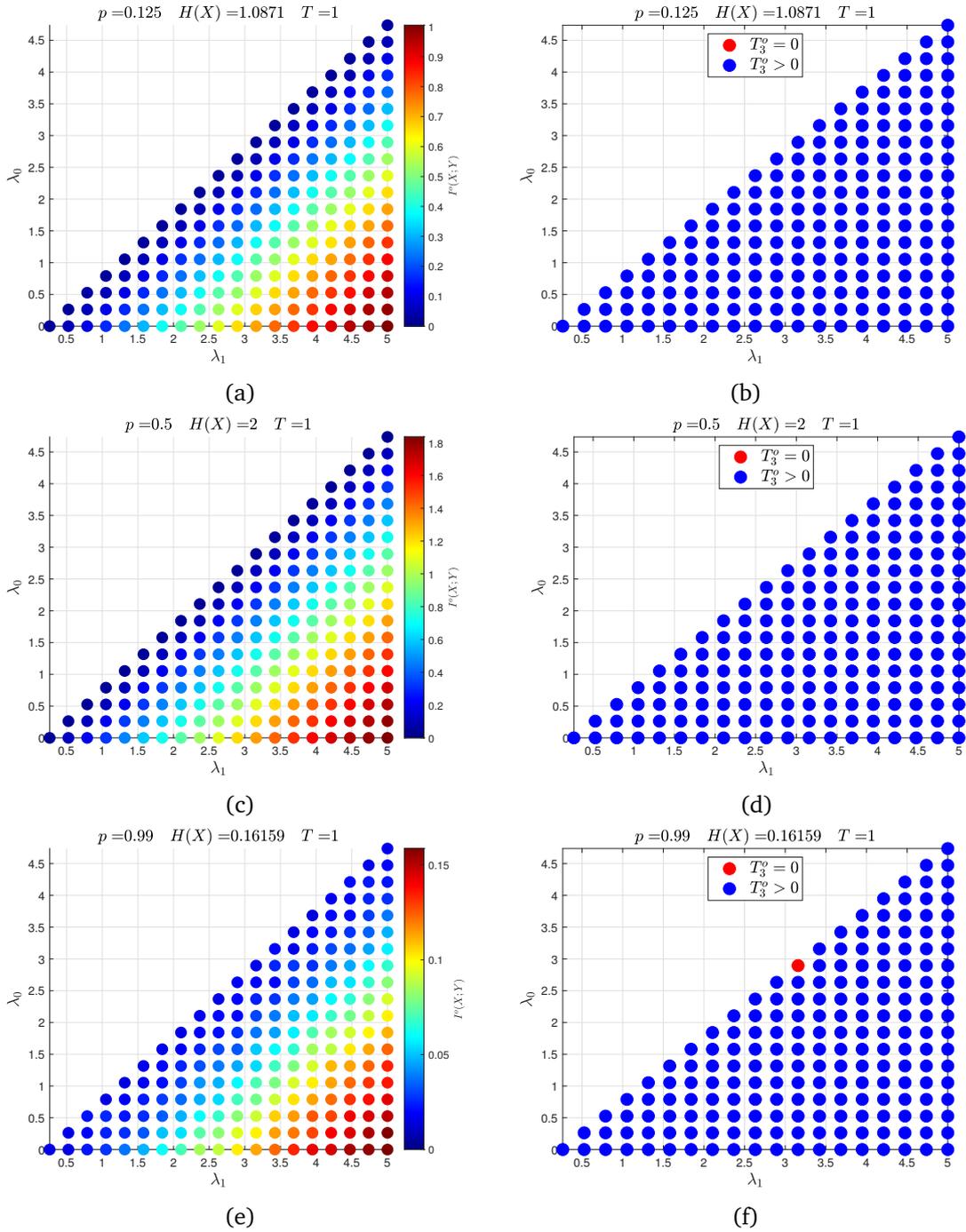


Fig. 6: Left: $I(X;Y)^O$ vs. (λ_0, λ_1) in the region $\lambda_1 > \lambda_0$, right: corresponding optimal argument parameter T_3^o vs. (λ_0, λ_1) for varying prior probabilities p . The search for each optimal argument T_3^o for any fixed: (λ_0, λ_1) and p is performed over the line $(T_1, T_2, T_3) := (\frac{1-T_3}{2}, \frac{1-T_3}{2}, T_3)$ where $0 \leq T_3 \leq 1$.

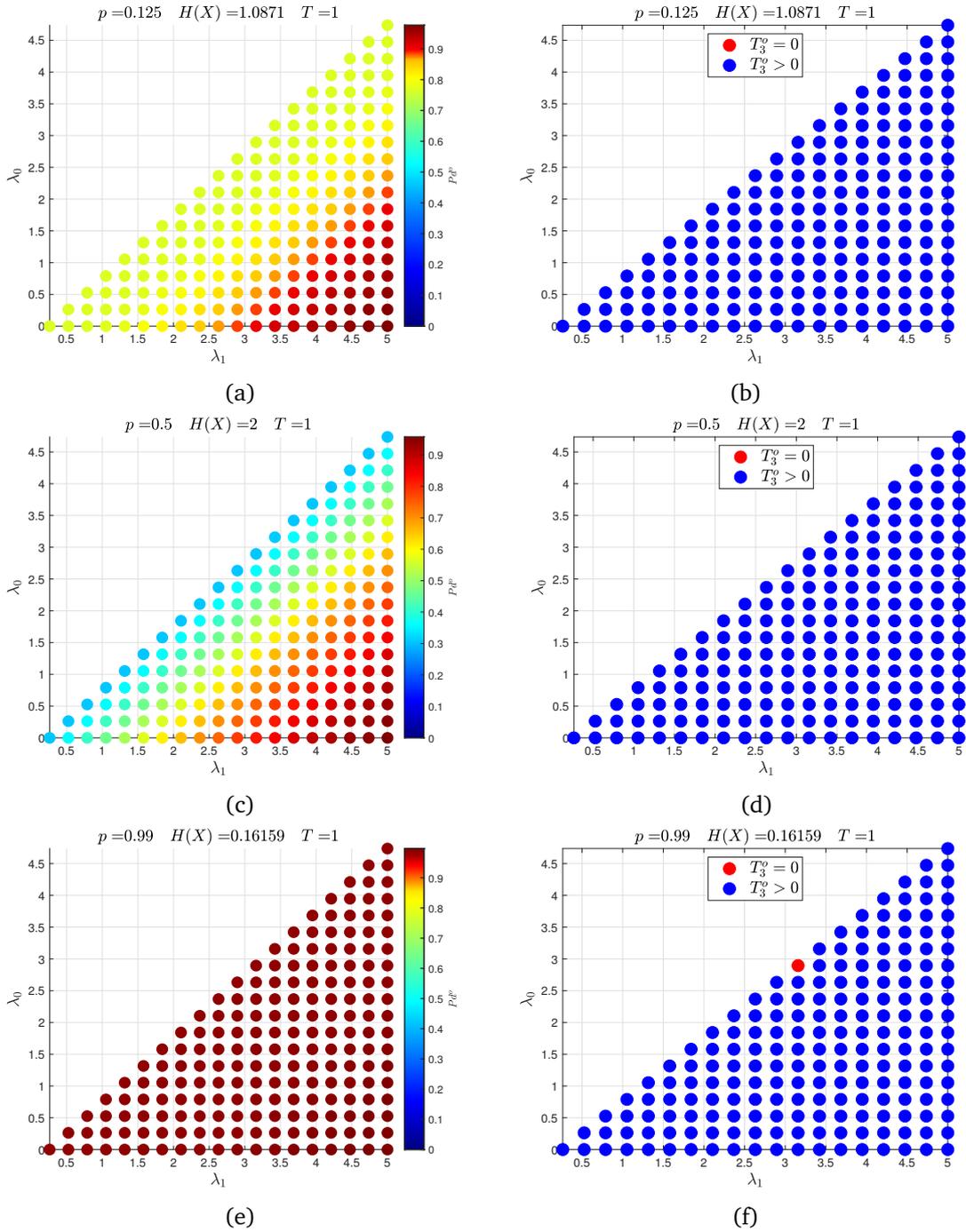


Fig. 7: Left: P_d^O vs. (λ_0, λ_1) in the region $\lambda_1 > \lambda_0$, right: corresponding optimal argument parameter T_3^o vs. (λ_0, λ_1) for varying prior probabilities p . The search for each optimal argument T_3^o for any fixed: (λ_0, λ_1) and p is performed over the line $(T_1, T_2, T_3) := (\frac{1-T_3}{2}, \frac{1-T_3}{2}, T_3)$ where $0 \leq T_3 \leq 1$.

6. CONCLUSION

This work attempts to address the problem of sensor scheduling in a vector Gaussian channel for a two target detection, when a specified structure on scaling matrix is imposed, using criteria of mutual information and Bayesian risk with 0 – 1 loss function. From computations, it was found that what is optimal argument under mutual information criterion need not necessarily be optimal under Bayesian inference. It was further found that mutual information is concave in the line $(T_1, T_2, T_3) := (\frac{T-\alpha}{2}, \frac{T-\alpha}{2}, \alpha)$ parametrized by $0 \leq \alpha \leq T$. For any given prior p and given finite time T : hybrid sensing is found to be the optimal sensing mechanism for any given time proportions.

There are few open questions that are worth exploring. For example, one can aim to find any counter-example in which mutual information is not Schur concave in the line $(T_1, T_2, T_3) := (\alpha \cdot T, (1 - \alpha) \cdot T, c)$ parametrized by $0 \leq \alpha \leq 1$ for some positive fixed constant c . One may also consider a problem where both the covariance matrix and the mean of the Gaussian channel are affected by the sensing time proportions and observe which sensing method is suitable for this case.

REFERENCES

- [1] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, 2005, doi: <https://doi.org/10.1109/tit.2005.844072>.
- [2] R. Atar and T. Weissman, "Mutual information, relative entropy, and estimation in the Poisson channel," *IEEE Transactions on Information Theory*, vol. 58, no. 3, pp. 1302–1318, 2012, doi: <https://doi.org/10.1109/tit.2011.2172572>.
- [3] M. Fahad and D. Fuhrmann, "Sensing Method for Two-Target Detection in Time-Constrained Vector Poisson Channel," *Signal & Image Processing : An International Journal (SIPIJ)*, vol. 12, no. 6, 2021, doi: <https://doi.org/10.5121/sipij.2021.12601>.
- [4] M. Fahad, "Sensing Methods for Two-Target and Four-Target Detection in Time-Constrained Vector Poisson and Gaussian Channels," Ph.D. dissertation, Michigan Technological University, Houghton, MI, 49931-1295, May 2021, doi: <https://doi.org/10.37099/mtu.dc.etr/1193>.
- [5] L. Wang, D. E. Carlson, M. R. Rodrigues, R. Calderbank, and L. Carin, "A Bregman matrix and the gradient of mutual information for vector Poisson and Gaussian channels," *IEEE Transactions on Information Theory*, vol. 60, no. 5, pp. 2611–2629, 2014, doi: <https://doi.org/10.1109/tit.2014.2307068>.
- [6] D. P. Palomar and S. Verdú, "Gradient of mutual information in linear vector Gaussian channels," *IEEE Transactions on Information Theory*, vol. 52, no. 1, pp. 141–154, 2006, doi: <https://doi.org/10.1109/tit.2005.860424>.
- [7] Y. Yang and R. S. Blum, "MIMO radar waveform design based on mutual information and minimum mean-square error estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 1, 2007, doi: <https://doi.org/10.1109/taes.2007.357137>.
- [8] S. Verdú, "Empirical Estimation of Information Measures: A Literature Guide," *Entropy*, vol. 21, no. 8, 2019. [Online]. Available: <https://www.mdpi.com/1099-4300/21/8/720>
- [9] M. Payaró and D. P. Palomar, "Hessian and concavity of mutual information, differential entropy, and entropy power in linear vector Gaussian channels," *IEEE Transactions on Information Theory*, vol. 55, no. 8, pp. 3613–3628, 2009, doi: <https://doi.org/10.1109/tit.2009.2023749>.
- [10] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004, doi: <https://doi.org/10.1017/cbo9780511804441>.
- [11] T. A. Schonhoff and A. A. Giordano, *Detection and Estimation Theory and its Applications*. Pearson College Division, 2006.



Muhammad Fahad (M'17) received the Ph.D. and M.S degrees in Electrical Engineering from Michigan Technological University, Houghton, MI in 2021 and 2016, respectively. He received M.S. degree in Electrical Engineering, from University of Engineering and Technology, Lahore, Pakistan and B.E. degree in Electronics Engineering from Mehran University of Engineering and Technology, Pakistan, in 2007 and 2005, respectively.

He was a Fulbright Ph.D. Scholar during 2011-2017 and currently serving as an Adjunct Assistant Professor in the Department of Applied Computing, College of Computing at Michigan Technological University, Houghton, MI. His research interests are in sensor management and Poisson processes.



Daniel R. Fuhrmann (S'78-M'84-SM'95-F'10) received the B.S.E.E. degree (*cum laude*) from Washington University, St. Louis, MO, USA, in 1979 and the M.A., M.S.E., and the Ph.D. degrees from Princeton University, Princeton, NJ, in 1982 and 1984, respectively.

From 1984 to 2008, he was with the Department of Electrical Engineering, now the Department of Electrical and Systems Engineering, Washington University. In the 2000-2001 academic year, he was a Fulbright Scholar visiting the Universidad Nacional de La Plata, Buenos Aires, Argentina. He is currently the Dave House Professor and Chair of the Department of Applied Computing, Michigan Technological University, Houghton, MI. His research interests lie in various areas of statistical signal and image processing,

including sensor array signal processing, radar systems, and adaptive sensing.

Dr. Fuhrmann is a former Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He was the Technical Program Chairman for the 1998 IEEE Signal Processing Workshop on Statistical Signal and Array Processing, and General Chairman of the 2003 IEEE Workshop on Statistical Signal Processing. He has been an ASEE Summer Faculty Research Fellow with the Naval Underwater Systems Center, New London, CT, a consultant to MIT Lincoln Laboratory, Lexington, MA, and an ASEE Summer Faculty Fellow with the Air Force Research Laboratory, Dayton, OH.