# SEQUENTIAL HYPOTHESIS TESTING WITH OFF-LINE RANDOMIZED SENSOR SELECTION STRATEGY

*Cheng-Zong Bai, Vijay Gupta and Yih-Fang Huang*

Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556, USA.
{cbai, vgupta2, huang}@nd.edu

**ABSTRACT**

We purpose and analyze an off-line randomized sensor selection strategy for sequential hypothesis testing problem constrained with sensor measurement costs. Within the framework of Wald's approximation, the sequential probability ratio test (SPRT) with sensor selection is designed for minimizing the expected total measurement cost subject to reliability and sensor usage constraints. In the case of symmetric hypotheses, we introduce a quantity, called efficiency, of a sensor and show that it is critical to the sensor selection in SPRT. Furthermore, an algorithm with linear time complexity is proposed to obtain the optimal sensor selection probabilities.

***Index Terms***— Sequential analysis, sequential probability ratio test, SPRT, sensor selection, linear fractional programming.

## 1. INTRODUCTION

Hypothesis testing when any of the multiple sensors present can be queried for new data is a basic function in a sensor network. However, this function is constrained by power consumption of sensors in a wireless sensor network. Thus, sensor selection for sequential hypothesis testing under sensor usage constraints is an important problem.

Sequential hypothesis testing has been considered since Wald's pioneering work on sequential analysis [1] which introduced the idea of sequential probability ratio test (SPRT) for the binary hypothesis testing problem and showed that the average sample number of SPRT can never be larger than any other sequential test with the same performance. Wald also gave several approximate formulae that make SPRT realizable. SPRT has since been extended in multiple directions. Particularly relevant to sensor networks is the decentralized SPRT studied by Veeravalli *et al.* [2][3] and Hussain [4]. Moreover, the problem of sensor selection for SPRT has been considered before. Srivastava *et al.* [5] propose randomized sensor selection rules for sequential testing of multiple hypotheses under several criteria. Polyanskiy *et al.* [6] focus on the binary hypothesis testing with a feedback controller. However, both of their works are based on asymptotic analysis. The sensor selection rule with finite sensor usage constraints can not be obtained by their approaches.

We consider a set-up in which several sensors are controlled by a fusion center. The fusion center applies a sequential decision rule for a binary hypothesis testing problem. At each time step, the fusion center assigns a sensor based on a sensor selection rule to take a measurement. We propose a stochastic sensor selection rule, in particular, an independent, identically distributed (i.i.d.) randomized off-line selection strategy that is analytically tractable. We call the proposed scheme an *off-line* selection scheme because the probability of selecting each sensor is preassigned and not updated. In practice, however, different sensors may take samples with different statistical characteristics. For example, the signal received by different sensors may have different signal-to-noise ratios (SNR). Our objective is to design a sequential test and a sensor selection rule that minimize the overall measurement cost (which is due to *e.g.,* making a measurement and the power required for data transmission to the fusion center) subject to certain reliability and sensor usage constraints.

This work is supported partly by AFOSR grant No. FA9550-10-1-0348 for the first and second coauthors.

## 2. MATHEMATICAL FORMULATION

Suppose that a binary, parametric, simple hypothesis $\{H_0, H_1\}$ whose realizations are observed by $J$ sensors in a sensor network. Conditioned on each hypothesis, we assume that the samples taken by a sensor are conditionally i.i.d., and the observations are conditionally independent from one sensor to another. Let the random variable $X_n$ represent the sample collected by the fusion center at the $n$-th time step. Denote the probability density function (pdf) of $X_n$, conditioned on $H_i$ with sensor $j$ being selected, by $f_i^j(x)$, which is assumed to be perfectly known to the fusion center. Define $\{S_n\}_{n=1}^{\infty}$ to be a random sequence where the event $\{S_n = j\}$ indicates that sensor $j$ is selected at time step $n$. For the off-line sensor selection strategy, $\{S_n\}_{n=1}^{\infty}$ is an i.i.d. random sequence with $P[S_n = j] = p_j$ and $\sum_{j=1}^{J} p_j = 1$. Also, let $0 < c_j < \infty$ be the cost of sensor $j$ taking a measurement and $\{C_n\}_{n=1}^{\infty}$ be a random sequence of the measurement cost at each time step. Obviously, $C_n = c_j$ if $S_n = j$. Thus $\{C_n\}_{n=1}^{\infty}$ is an i.i.d. random sequence as well.

The expected total measurement cost of the sequential test can be expressed as $\mathbb{E}[\sum_{n=1}^{N} C_n]$ where $N$, a random variable, is the stopping time of the sequential test. By Wald's identity [7], $\mathbb{E}[\sum_{n=1}^{N} C_n]$ is the product of $\mathbb{E}[C_n]$ and $\mathbb{E}[N]$, namely,

$$\mathbb{E}\left[\sum_{n=1}^{N} C_n\right] = \mathbb{E}[C_n]\,\mathbb{E}[N]. \tag{1}$$

Moreover, let $N_j$ be the total number of samples taken by sensor $j$. From Wald's identity, we can show that $\mathbb{E}[N_j] = p_j\mathbb{E}[N]$ by constructing an i.i.d. Bernoulli random sequence $\{Y_n\}_{n=1}^{\infty}$ such that $Y_n = 1$ if $S_n = j$ and zero otherwise.

Now the fusion center performs a sequential hypothesis test with an off-line sensor selection strategy that minimizes the expected total measurement cost $\mathbb{E}[C_n]\,\mathbb{E}[N]$. We define $\alpha \leq \alpha_0, \beta \geq \beta_0$ as the reliability constraints of the test, where $\alpha$, $\beta$ are the probability of false alarm and probability of detection, respectively. Furthermore, there is an additional set of constraints of the usage of individual sensors that $\mathbb{E}[N_j] \leq n_j, n_j > 0, j = 1, \ldots, J$, motivated by the limitation of the battery life of a sensor. We observe that, however, $\mathbb{E}[C_n]$ is not a function of the stopping rule and terminal decision rule of the sequential test. By the Wald-Wolfowitz theorem [1], SPRT is the optimal sequential test that minimizes $\mathbb{E}[N]$ subject to the constraints, which implies that an SPRT$(a, b)$ remains optimal in this scenario.

## 3. SPRT WITH OFF-LINE RANDOMIZED SENSOR SELECTION STRATEGY

Suppose that sensor $j$ is selected at time step $n$, the log-likelihood ratio (LLR) of $X_n$ is given by $L(X_n|S_n = j) = \log[f_1^j(X_n)/f_0^j(X_n)]$. The LLR of $\boldsymbol{X}_1^n = \{X_1, \ldots, X_n\}$, conditioned on the sensor selection pattern $\boldsymbol{S}_1^n = \boldsymbol{s}_1^n$, is the sum of individual LLR, *i.e.*, $L(\boldsymbol{X}_1^n|\boldsymbol{S}_1^n = \boldsymbol{s}_1^n) = \sum_{i=1}^n L(X_i|S_i = s_i)$, where $\boldsymbol{s}_1^n = \{s_1, \ldots, s_n\}$. Thus the test is of the following form: continue sampling if $a < L(\boldsymbol{X}_1^n|\boldsymbol{S}_1^n = \boldsymbol{s}_1^n) < b$, otherwise, terminate and choose a hypothesis.

The exact values of the thresholds $a, b$ for given $\alpha_0, \beta_0$ seem intractable in general. Fortunately, Wald's approximations [1] can be easily generalized to suit this case. The approximate thresholds are given by

$$a \cong \log\left(\frac{1-\beta_0}{1-\alpha_0}\right) \text{ and } b \cong \log\left(\frac{\beta_0}{\alpha_0}\right). \tag{2}$$

Notably, the approximate values of $a, b$ are not functions of $\boldsymbol{p} = [p_1, \ldots, p_J]^T$, which implies that, under Wald's approximations, the jointly optimization problem with respect to the triplet $(a, b, \boldsymbol{p})$ can be separated into two parts, *i.e.*,

$$\min_{\boldsymbol{p}} \left\{ \boldsymbol{c}^T \boldsymbol{p} \times \left\{ \min_{\{a,b\}}, \mathbb{E}[N] \right\} \right\}, \tag{3}$$

subject to $\alpha \leq \alpha_0, \beta \geq \beta_0$ and $\mathbb{E}[N_j] \leq n_j, j = 1, \ldots, J$, where $\boldsymbol{c} = [c_1, \ldots, J]^T$, and the inner minimization is approximately solved by (2). Define $\boldsymbol{d}_i = [d_i^1, \ldots, d_i^J]^T$, $i \in \{0, 1\}$, where $d_0^j = -D(f_0^j||f_1^j)$, $d_1^j = D(f_1^j||f_0^j)$, and $D(f||g)$ denotes the Kullback-Leibler divergence between the pdf $f$ and $g$. Because $\{L(X_n|S_n)\}_{n=1}^\infty$ is a conditional i.i.d. random sequence, by Bayes rule, Wald's identity and Wald's approximations, an approximate average sample number can be obtained by

$$\begin{aligned}\mathbb{E}[N] &= (1-\pi_1)\frac{\mathbb{E}\left[L\left(\boldsymbol{X}_1^N|\boldsymbol{S}_1^N\right)|H_0\right]}{\mathbb{E}\left[L\left(X_n|S_n\right)|H_0\right]} + \pi_1\frac{\mathbb{E}\left[L\left(\boldsymbol{X}_1^N|\boldsymbol{S}_1^N\right)|H_1\right]}{\mathbb{E}\left[L\left(X_n|S_n\right)|H_1\right]} \\ &\cong \frac{(1-\pi_1)A_0}{\boldsymbol{d}_0^T\boldsymbol{p}} + \frac{\pi_1 A_1}{\boldsymbol{d}_1^T\boldsymbol{p}}, \end{aligned} \tag{4}$$

where $\pi_1$ is the *a priori* probability of $H_1$, $A_0 = (1-\alpha_0)a + \alpha_0 b$, $A_1 = (1-\beta_0)a + \beta_0 b$. Here we have assumed that $-\infty < d_0^j < 0$, $0 < d_1^j < \infty, \forall j \in \{1, \ldots, J\}$ and $A_0 < 0, A_1 > 0$.

## 4. THE CASE OF SYMMETRIC HYPOTHESES

For the rest of the paper, we concentrate on the class of binary hypothesis with symmetric Kullback-Leibler divergence. As an example, consider the following binary hypothesis testing problem,

$$H_0 : f_0^j(x_n) = \mathcal{N}(0, \sigma_j^2) \text{ versus } H_1 : f_1^j(x_n) = \mathcal{N}(\theta_j, \sigma_j^2), \tag{5}$$

where $\mathcal{N}(\theta, \sigma^2)$ denotes to a Gaussian pdf with mean $\theta$ and variance $\sigma^2$. This problem arises in, *e.g.*, the detection of amplitude in an additive white Gaussian noise channel. In this problem, $d_0^j = -\theta_j^2/(2\sigma_j^2)$ and $d_1^j = \theta_j^2/(2\sigma_j^2)$. If the condition $\boldsymbol{d}_1 = -\boldsymbol{d}_0$ is satisfied, we say that the hypotheses are symmetric. In this case, from (4), the minimization problem in (3) can be simplified to

$$\begin{aligned} &\min_{\boldsymbol{p}} f(\boldsymbol{p}) = \frac{\boldsymbol{c}^T\boldsymbol{p}}{\boldsymbol{d}^T\boldsymbol{p}} \\ &\text{Subject to } \boldsymbol{d}^T\boldsymbol{p} - \frac{A}{n_j}p_j \geq 0, j = 1, \ldots, J \\ &\boldsymbol{p} \in \mathcal{P}_J \end{aligned} \tag{6}$$

where $\boldsymbol{d} = [d_1, \ldots, d_J]^T = \boldsymbol{d}_1$, $A = -(1-\pi_1)A_0 + \pi_1 A_1 > 0$ and $\mathcal{P}_J$ denotes to the probability simplex in $\mathbb{R}^J$. In (6), the objective function is a linear fractional function, and the feasible region is a polyhedron in $\mathbb{R}^J$. Thus the optimization problem for the symmetric hypotheses is a linear fractional programming (LFP). It is known that an LFP can be converted to a linear programming problem [8]. Hence, (6) can be solved by, for example, simplex method or interior point method. However, because of the special structure of our LFP problem, we propose a computationally efficient algorithm for solving (6) in the subsequent subsection.

### 4.1. Sufficient Conditions of an Optimal Solution to the LFP

If the sensor network consists of only a single sensor $j$, the average total measurement cost of the single sensor SPRT is $Ac_j/d_j$. Therefore the ratio $c_j/d_j$ plays an important role in SPRT. To generalize this concept, we define the efficiency of sensor $j$ to be $e_j = d_j/c_j$. Accordingly, we say that sensor $i$ is more efficient than sensor $j$ if $e_i > e_j$. Finally, the sensors are said to be degenerate if they have the same efficiency. Without loss of generality, we assume that the indices of the sensors are ordered by their efficiency, *i.e.,*

$$e_1 \geq e_2 \geq \ldots \geq e_J. \tag{7}$$

We next develop sufficient conditions for an optimal solution based on the following four lemmas.

**Lemma 1.** *If $\mathcal{C} \neq \emptyset$, there exists a vertex point of $\mathcal{C}$ which is an optimal solution to the LFP in (6), where $\mathcal{C}$ denotes to the feasible region in (6).*

*Proof.* Since $\mathcal{C}$ is bounded, this lemma follows easily from the quasiconvexity of linear fractional functions [9]. □

It is well known that linear inequality and equality constraints can be converted to linear constraints in the standard form considered in linear programming by introducing slack variables. By Lemma 1, there is an optimal solution occurring at a vertex point of $\mathcal{C}$. In other words, there must be a basic feasible solution to the standard form that is optimal. The following lemma provides a way to determine those basic variables.

**Lemma 2.** *If $\boldsymbol{p} = [p_1, \ldots, p_J]^T$ is a minimum feasible solution to (6) and there exists $j$, where $0 < j \leq J, p_j > 0$, then for every $k \in \{i : e_i > e_j\}$, the equation $\boldsymbol{d}^T\boldsymbol{p} - Ap_k/n_k = 0$ holds.*

*Proof.* Assume that $\boldsymbol{p} = [p_1, \ldots, p_J]^T$ is a minimum feasible solution and there exists $k \in \{i : e_i > e_j\}$ such that $\boldsymbol{d}^T\boldsymbol{p} - Dp_k/n_k > 0$. Let $\boldsymbol{p}' = [p_1', \ldots, p_J']^T$, where $p_j' = (p_j - \varepsilon_1)/\zeta$, $p_k' = (p_k + \varepsilon_2)/\zeta$, $\varepsilon_1, \varepsilon_2 > 0$ and $p_i' = p_i/\zeta, i \neq j, k$. $\zeta$ is the normalization constant that guarantees $\sum_{i=1}^J p_i = 1$ and it is given by $\zeta = 1 + \varepsilon_2 - \varepsilon_1$. We further let $\boldsymbol{d}^T\boldsymbol{p}' = \boldsymbol{d}^T\boldsymbol{p}/\zeta$, then we have

$$\boldsymbol{d}^T\boldsymbol{p}' = \frac{1}{\zeta}(\boldsymbol{d}^T\boldsymbol{p} - \varepsilon_1 d_j + \varepsilon_2 d_k) = \frac{1}{\zeta}\boldsymbol{d}^T\boldsymbol{p}$$

which yields

$$\varepsilon_1 d_j = \varepsilon_2 d_k. \tag{8}$$

Next, we need to check the feasibility of $\boldsymbol{p}'$.

$$\frac{Ap_i'}{\boldsymbol{d}^T\boldsymbol{p}'} = \frac{A(p_i/\zeta)}{\boldsymbol{d}^T\boldsymbol{p}/\zeta} = \frac{Ap_i}{\boldsymbol{d}^T\boldsymbol{p}} \le n_i, (i \ne j, k) \tag{9}$$

$$\frac{Ap_j'}{\boldsymbol{d}^T\boldsymbol{p}'} = \frac{p_j - \varepsilon_1}{\zeta}\frac{A}{\boldsymbol{d}^T\boldsymbol{p}/\zeta} = \frac{Ap_j}{\boldsymbol{d}^T\boldsymbol{p}} - \frac{A\varepsilon_1}{\boldsymbol{d}^T\boldsymbol{p}} < n_j \tag{10}$$

$$\frac{Ap_k'}{\boldsymbol{d}^T\boldsymbol{p}'} = \frac{p_k + \varepsilon_2}{\zeta}\frac{A}{\boldsymbol{d}^T\boldsymbol{p}/\zeta} = \frac{Ap_k}{\boldsymbol{d}^T\boldsymbol{p}} + \frac{A\varepsilon_2}{\boldsymbol{d}^T\boldsymbol{p}} \tag{11}$$

Since $\boldsymbol{d}^T\boldsymbol{p} - Ap_k/n_k > 0$, we can always select a sufficient small $\varepsilon_2$ such that (11) is still less or equal to $n_k$ and $p_j \ge 0$. Consequently, from (9), (10), (11), $\boldsymbol{p}'$ is feasible. Finally, by (8),

$$\begin{aligned} f(\boldsymbol{p}') &= \frac{(\boldsymbol{c}^T\boldsymbol{p} - c_j\varepsilon_1 + c_k\varepsilon_2)/\zeta}{\boldsymbol{d}^T\boldsymbol{p}/\zeta} = \frac{\boldsymbol{c}^T\boldsymbol{p} - \varepsilon_2 d_k\left(e_j^{-1} - e_k^{-1}\right)}{\boldsymbol{\mu}^T\boldsymbol{p}} \\ &< \frac{\boldsymbol{c}^T\boldsymbol{p}}{\boldsymbol{d}^T\boldsymbol{p}} = f(\boldsymbol{p}) \end{aligned}$$

which contradicts to the assumption that $\boldsymbol{p}$ is a minimum solution. Hence the statement holds. □

Lemma 2 indicates that, for an optimal $\boldsymbol{p}$, if $p_j > 0$ is a basic variable, then every $p_k$ is also a basic variable, where sensor $k$ is more efficient than sensor $j$, and the slack variable associated with $\boldsymbol{d}^T\boldsymbol{p} - Ap_k/n_k \ge 0$ must be nonbasic. In other words, we must never skip a more efficient sensor to use a less efficient one. Moreover, we should continue using an efficient sensor until it meets the equality of $\mathbb{E}[N_j] \le n_j$. As a result, from Lemma 2, an optimal solution to (6) is in the form of

$$\boldsymbol{p} = [p_1, \ldots, p_j, p_{j+1}, \ldots, p_{j+l}, \boldsymbol{0}^T_{J-(j+l)}]^T, \tag{12}$$

that satisfies $\boldsymbol{d}^T\boldsymbol{p} - Ap_k/n_k = 0, k = 1, \ldots, j$ where $e_j > e_{j+1} = \ldots = e_{j+l}$, and $\boldsymbol{0}_i$ denotes to an $(i \times 1)$ zero vector. However, Lemma 2 does not show how to determine the basic variables for the $l$ degenerate sensors. Lemma 3 will show that there is no difference between those degenerate sensors.

**Lemma 3.** *Every feasible $\boldsymbol{p}$ in the form of (12) that satisfies the equations $\boldsymbol{d}^T\boldsymbol{p} - Ap_k/n_k = 0, k = 1, \ldots, j$, has the same value of $f(\boldsymbol{p})$.*

*Proof.* Note that $p_k = n_k\boldsymbol{d}^T\boldsymbol{p}/A, k = 1, \ldots, j$. Then we have

$$\begin{aligned} f(\boldsymbol{p}) &= \frac{\sum_{i=1}^{j} c_i p_i + \sum_{i=j+1}^{j+l} c_i p_i}{\boldsymbol{\mu}^T\boldsymbol{p}} \\ &= \frac{\sum_{i=1}^{j} c_i p_i + e_{j+1}^{-1}\sum_{i=j+1}^{j+l} d_i p_i}{\boldsymbol{d}^T\boldsymbol{p}} \\ &= \frac{\sum_{i=1}^{j} c_i p_i + e_{j+1}^{-1}(\boldsymbol{d}^T\boldsymbol{p} - \sum_{i=1}^{j} d_i p_i)}{\boldsymbol{d}^T\boldsymbol{p}} \\ &= \frac{\sum_{i=1}^{j} c_i n_i \boldsymbol{d}^T\boldsymbol{p}/A + e_{j+1}^{-1}(\boldsymbol{d}^T\boldsymbol{p} - \sum_{i=1}^{j} d_i n_i \boldsymbol{d}^T\boldsymbol{p}/A)}{\boldsymbol{d}^T\boldsymbol{p}} \\ &= e_{j+1}^{-1} - \sum_{i=1}^{j}\frac{d_i n_i}{A}\left(e_i^{-1} - e_{j+1}^{-1}\right) \end{aligned} \tag{13}$$

which is a constant. □

**Lemma 4.** *Let $j, k$ be two indices such that $e_j > e_k$. Define $\boldsymbol{p} = [p_1, \ldots, p_j, \boldsymbol{0}^T_{J-j}]^T$ where $\boldsymbol{d}^T\boldsymbol{p} - Ap_i/p_i = 0, i = 1, \ldots, j-1$ and $\boldsymbol{p}' = [p_1', \ldots, p_k', \boldsymbol{0}^T_{J-k}]^T$ where $\boldsymbol{d}^T\boldsymbol{p}' - Ap_i'/n_i = 0, i = 1, \ldots, k-1$. Then $f(\boldsymbol{p}) < f(\boldsymbol{p}')$.*

*Proof.* We rewrite $f(\boldsymbol{p})$ and $f(\boldsymbol{p}')$ by (13), then we obtain

$$\begin{aligned} &f(\boldsymbol{p}) - f(\boldsymbol{p}') \\ &= \left(e_j^{-1} - e_k^{-1}\right) - \sum_{i=1}^{j-1}\frac{d_i n_i}{A}\left(e_i^{-1} - e_j^{-1}\right) + \sum_{i=1}^{k-1}\frac{d_i n_i}{A}\left(e_i^{-1} - e_k^{-1}\right) \\ &= \left(e_j^{-1} - e_k^{-1}\right)\left(1 + \sum_{i=1}^{j-1}\frac{d_i n_i}{A}\right) + \sum_{i=j}^{k-1}\frac{d_i n_i}{A}\left(e_i^{-1} - e_k^{-1}\right) < 0 \end{aligned}$$

□

Using Lemma 1, 2, 3 and 4, we can now state the main result of the section.

**Theorem 1.** *If a vector $\boldsymbol{p}^{(j)} \in \mathbb{R}^J$ satisfies all the following three conditions, then it is a minimum solution to the LFP in (6).*

1. *$\boldsymbol{p}^{(j)}$ is in the form of $\boldsymbol{p}^{(j)} = [p_1^{(j)}, \ldots, p_j^{(j)}, \boldsymbol{0}^T_{J-j}]^T$ where $\boldsymbol{d}^T\boldsymbol{p}^{(j)} - Ap_i^{(j)}/n_i = 0$, $i = 1, \ldots, j-1$.*
2. *$\boldsymbol{p}^{(j)} \in \mathcal{C}$.*
3. *There exists no $k < j$ such that $\boldsymbol{p}^{(k)}$ is feasible where $\boldsymbol{p}^{(k)}$ is in the same form of the first condition.*

### 4.2. An Efficient Algorithm for Solving the LFP

By Theorem 1, an optimal solution to the LFP can be found by the following algorithm. Initially, we start by selecting the most efficient sensor with probability one. At the $j$-th iteration, we select the $j$ most efficient sensors and solve $\boldsymbol{p}^{(j)}$ from the equations $\boldsymbol{d}^T\boldsymbol{p}^{(j)} - Ap_i^{(j)}/n_i = 0, i = 1, \ldots, j-1$ and $\sum_{i=1}^{j} p_i^{(j)} = 1$. If it is feasible, it must be optimal and the algorithm is terminated. Otherwise we try to use $j+1$ sensors and repeat the same procedure until a feasible solution is obtained. If all $J$ sensors are included and $\boldsymbol{p}^{(J)}$ is still infeasible, it implies that the feasible region is an empty set. Nevertheless, at each iteration $j$, we need to solve a $j \times j$ system of linear equations. We note that $p_k^{(j)} = n_k/\mathbb{E}[N], k = 1, \ldots, j-1$. Thus the $j \times j$ system of linear equations can be reduced into two variables $\mathbb{E}[N]$, $p_j^{(j)}$ and two equations

$$\boldsymbol{d}^T\boldsymbol{p}^{(j)} = \sum_{i=1}^{j-1}\frac{d_i n_i}{\mathbb{E}[N]} + d_j p_j^{(j)} = \frac{A}{n_k}p_k^{(j)} = \frac{A}{\mathbb{E}[N]} \tag{14}$$

$$\sum_{i=1}^{j} p_i^{(j)} = \sum_{i=1}^{j-1}\frac{n_i}{\mathbb{E}[N]} + p_j^{(j)} = 1. \tag{15}$$

The solutions to (14), (15) are given by,

$$\mathbb{E}[N] = \frac{A - \Sigma_{dn}(j-1)}{d_j} + \Sigma_n(j-1), \tag{16}$$

$$p_j^{(j)} = 1 - \frac{\Sigma_n(j-1)}{\mathbb{E}[N]}, \tag{17}$$

where $\Sigma_n(j) = \sum_{i=1}^{j} n_i$ and $\Sigma_{dn}(j) = \sum_{i=1}^{j} d_i n_i$. Furthermore, the feasibility test of $\boldsymbol{p}^{(j)}$ consists of two parts, the tests of $p_i^{(j)} \ge 0$ and $\boldsymbol{d}^T\boldsymbol{p}^{(j)} - Ap_i^{(j)}/n_i \ge 0, i = 1, \ldots, j$. For the first part, we

**Table 1**. Numerical results of SPRT with optimal sensor selection

| | |
|---|---|
| $\boldsymbol{p}$ | $[0.257, 0.343, 0.214, 0.171, 0.015, 0, 0, 0]^T$ |
| $\mathbb{E}[N]$ | 23.36 (Wald's apx.), 24.48 (Sim.) |
| $\mathbb{E}[N_j]$ | $\{6, 8, 5, 4, 0.36, 0, 0, 0\}$ (Wald's apx.) |
| | $\{6.29, 8.39, 5.24, 4.19, 0.38, 0, 0, 0\}$ (Sim.) |
| $\mathbb{E}[\sum_{n=1}^{N} C(n)]$ | 55.76 (Wald's apx.), 58.45 (Sim.) |

**Table 2**. Numerical results of SPRT with equally likely selection

| | |
|---|---|
| $\boldsymbol{p}$ | $[0.125, 0.125, \ldots, 0.125]^T$ |
| $\mathbb{E}[N]$ | 29.14 (Wald's apx.), 30.39 (Sim.) |
| $\mathbb{E}[N_j]$ | $\{3.64, 3.64, \ldots, 3.64\}$ (Wald's apx.) |
| | $\{3.80, 3.80, \ldots, 3.80\}$ (Sim.) |
| $\mathbb{E}[\sum_{n=1}^{N} C(n)]$ | 65.09 (Wald's apx.), 67.88 (Sim.) |

only need to check $\mathbb{E}[N] > 0$ and $p_j \geq 0$. By (14), $\mathbb{E}[N]$ can be alternatively represented by $\mathbb{E}[N] = [A - \Sigma_{dn}(j-1)]/(d_j p_j^{(j)})$. Then we have $p_i^{(j)} \geq 0$, $i = 1, \ldots, j$, if and only if $A - \Sigma_{dn}(j-1) \geq 0$ and $p_j^{(j)} > 0$. For the second part, since $\boldsymbol{d}^T \boldsymbol{p}^{(j)} - A p_i^{(j)}/n_i = 0$, $i = 1, \ldots, j-1$, always satisfies the constraint, we only need to examine whether $\mathbb{E}[N_j] = p_j^{(j)} \mathbb{E}[N] \leq n_j$ or not. Hence the solution to the LFP in (6) can be iteratively solved by the efficient algorithm as follows.

**Algorithm**

Order the indices of the sensors such that (7) is satisfied.
$\Sigma_{dn} \leftarrow 0$
$\Sigma_n \leftarrow 0$
**if** $A/d_1 \leq n_1$ **then**
  $\boldsymbol{p}^{(1)} \leftarrow [1, \mathbf{0}_{J-1}^T]^T$
  **return** $\boldsymbol{p}^{(1)}$
**else**
  **for** $j = 2 \rightarrow J$ **do**
    $\Sigma_{dn} \leftarrow \Sigma_{dn} + d_{j-1} n_{j-1}$
    $\Sigma_n \leftarrow \Sigma_n + n_{j-1}$
    **if** $A - \Sigma_{dn} \geq 0$ **then**
      $\mathbb{E}[N] \leftarrow (A - \Sigma_{dn})/d_j + \Sigma_n$
      $p_j^{(j)} \leftarrow 1 - \Sigma_n / \mathbb{E}[N]$
      **if** $p_j^{(j)} > 0$ and $p_j^{(j)} \mathbb{E}[N] \leq n_j$
        $\boldsymbol{p}^{(j)} \leftarrow [n_1/\mathbb{E}[N], \ldots, n_{j-1}/\mathbb{E}[N], p_j^{(j)}, \mathbf{0}_{J-j}^T]^T$
        **return** $\boldsymbol{p}^{(j)}$
      **end if**
    **end if**
  **end for**
**end if**

If the indices have been ordered, the time complexity of the algorithm is of $\mathcal{O}(J)$ which is much more computationally efficient than general algorithms of linear programming.

## 5. NUMERICAL RESULTS

For the hypothesis testing problem given in (5), we define the SNR of the sample of sensor $j$ to be $\text{SNR}_j = \theta_j^2/\sigma_j^2$. Let $J = 8$ and the SNR's be $\{3.5, 3, 2.5, 2, 1.5, 1, 0.5, 0\}$ dB, respectively. Let the measurement cost be $c_j = 1 + \sqrt{\text{SNR}_j}$ where $\text{SNR}_j$ is in linear scale. Moreover, the constraints are set by $n_j = \{6, 8, 5, 4, 8, 4, 8, 6\}$. Suppose $\pi_1 = 0.2$, and we require $\alpha \leq \alpha_0 = 10^{-9}$, $\beta \geq \beta_0 = 1 - 10^{-10}$. We perform SPRT with two different off-line sensor selection schemes, the sensor selection probabilities obtained by (6) and the equally likely selection strategy. Notice that the equally likely scheme is not always feasible for arbitrary constraints. The comparisons between the two schemes are summarized in Table 1 and 2. We present numerical results obtained by both simulations and Wald's approximation. Compared with the equally likely selection, around 14% of average measurement cost is reduced by adopting the optimal sensor selection probabilities. Nevertheless, there is a slight discrepancy between the simulations and analytical results due to the error of Wald's approximation. The discrepancy causes the violation of constraints since Wald's approximation underestimates the average sample number. A possible way to avoid the violation is to set a safety margin for $n_j$ by considering an upper bound of the actual value of $\mathbb{E}[N]$.

## 6. CONCLUDING REMARKS

Employing Wald's approximation, an SPRT with off-line sensor selection rule is presented. For the case of symmetric hypotheses, the problem is equivalent to an LFP. It is shown that an optimal solution can be obtained by making the number of relatively efficient sensors that meet their corresponding usage constraints as small as possible. An algorithm with linear computational complexity for solving the LFP is also proposed. Numerical results show that selecting sensors with our proposed optimal strategy will significantly reduce the average total measurement cost, compared with the equally likely selection scheme.

## 7. REFERENCES


[1] A. Wald, *Sequential Analysis*, Wiley, New York, 1947.

[2] V. V. Veeravalli, T. Basar and H. V. Poor, "Decentralized Sequential Detection with a Fusion Center Performing the Sequential Test," *IEEE Trans. on Information Theory*, 39(2): 433–442, March 1993.

[3] V. V. Veeravalli, T. Basar and H. V. Poor, "Decentralized Sequential Detection with Sensors Performing Sequential Tests," *Math. Control Signals Systems*, 7(4): 292-305, December 1994.

[4] A. M. Hussain, "Multisensor Distributed Sequential Detection," *IEEE Trans. on Aerospace and Electronic Systems*, 30(3), 698–708, July 1994.

[5] V. Srivastava, K. Plarre and F. Bullo, "Randomized Sensor Selection in Sequential Hypothesis Testing," *IEEE Trans. on Signal Processing*, 59(5): 2342–2354, May 2011.

[6] Y. Polyanskiy and S. Verdú, "Binary hypothesis testing with feedback (slides)," *2011 Information Theory and Applications Workshop (ITA)*, UCSD, San Diego, CA, USA, February 2011.

[7] D. Siegmund, *Sequential Analysis*, Springer-Verlag, New York, 1985.

[8] A. Charnes and W. W. Cooper, "Programming with Linear Fractional Functions," *Naval Research Logistics*, 9: 181–186, 1962.

[9] M. S. Bazaraa, H. D. Sherali and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 3rd ed., Wiely, New Jersey, 2006.