Optimal Sensor Selection for Binary Detection based on Stochastic Submodular Optimization

João P. Hespanha Denis Garagić

***Abstract*— We address the problem of selecting sensors for the estimation of binary random variables, so as to minimize the probability of error. This problem arises when a large number of sensors are potentially available, but only a few can actually be used for estimation purposes. While sensor selection is a combinatorial problem, we show that the optimization of an upper bound on the probability of error can be formulated as a submodular maximization for which computationally efficient algorithms can provide solutions with guaranteed performance. The submodular optimization that needs to be solved involves the computation of an expected value that generally cannot be computed in closed form, but we show that replacing the expected value by a Monte Carlo empirical mean can result in negligible loss of performance with high probability. We illustrate the use of these results in the context of detecting illegal unreported and unregulated (IUU) fishing.**

I. Introduction

Binary detection refers to the problem of making a binary decision based on noisy measurements, while minimizing the probability of error. In many such detection problems, part of the decision process involves selecting the set of sensors that will provide the measurements without knowing the actual values of the measurements that the sensors will report.

This paper addresses the optimal sensor selection for binary detection, in scenarios where the decision maker has a "budget" that limits how many sensors can be queried and wants to "purchase" measurements from those sensors that will result in the smallest probability of error. Here, "budget" and "purchase" should be understood in a broad sense. For example, when decisions are made based on very large datasets, "purchasing" a particular measurement may mean running a classification algorithm on a subset of the available dataset and the decision-maker's "budget" may correspond to a limitation in computational power.

Optimal sensor selection arises in numerous domains but our work is mainly motivated by problems in environmental conservation and computer security. The detection of illegal, unreported and unregulated (IUU) fishing is a national priority to many nations including the USA and recent reports on IUU fishing estimate that one in five fish in global markets are caught by vessels illegally, amounting to about $23.5B per year [3]. A wide range of measurements can be used for the detection of IUU fishing, ranging from the time series of the vessels' automatic identification system (AIS) to satellite imagery [16]. AIS data is widely available and inexpensive [1], but it can be manipulated by vessels engaged in IUU [6, 9] and it is hard to interpret, so the continuous monitoring of global AIS data is computationally very demanding. Satellite imagery can be much more reliable, but is expensive.

In computer security, sensor selection is generally needed to address computational limitations, which force cyber-defense systems to make decisions based just on a subset of the data that is available. Cyber-sensors that monitor low-level events like network traffic, email activity, access to services or files and user authentication generate volumes of data that cannot be extensively processed in real-time [12]. Instead, anomaly detection systems typically focus their efforts on small subsets of the full data logs.

We formulate the sensor selection problem as an optimization over a large set of L random variables, each representing one sensor measurement, that seeks to find the "best" $K \ll L$ random variables. In this context, "best" should be understood as getting the smallest detection error for an optimal estimator that uses these K variables (see Section II).

Optimal sensor selection problems are typically computationally difficult and the one considered here is no exception. However, we show that the function that maps sets of sensor random variables to the probability of error of the optimal detector can be upper-bounded in terms of a submodular function. Moreover, minimizing this upper bound is equivalent to a submodular maximization (see Section III), which enables the use of tools from submodular optimization in optimal sensor selection problems for binary detection.

A key difficulty in solving the submodular maximization arising in sensor selection is that the function to be maximized takes the form of an expected value over the sensor measurements, which is typically difficult to compute in closed form. To overcome this difficulty we use a stochastic version of the classical submodular greedy algorithm that makes selections based on an empirical expectation rather than the true expected value. A sample-complexity bound is provided in Section IV, which guarantees with high probability that the proposed stochastic greedy algorithm does not fall below $(1-e^{-1})$ of the global maximum by more than a predefined offset $\epsilon > 0$, which can be made arbitrarily small.

J. Hespanha is with the University of California, Santa Barbara, USA. D. Garagić is with BAE Systems, FAST Labs, Burlington, MA, USA.
This material is based upon work supported by DARPA under the contract HT0011990016, by the U.S. Office of Naval Research under the MURI grant No. N00014-16-1-2710, and by the Air Force Research Laboratory under the contract FA8750-18-C-0014. The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

The results in this paper are illustrated through numerical examples related to IUU fishing. Specifically, given a reported path for a vessel suspected of engaging in IUU fishing, we determine where one should gather satellite imaging and AIS tracks to provide the best estimate of whether or not the vessel is indeed engaging in fishing. The results shown are based on data made available by Global Fishing Watch [1], from which one can extract information about fishing effort (measured in hour per day) since 2012 for the whole Earth.

A. Related Work

Sensor selection is closely related to experiment design for which a large literature is available. While sensor selection is a combinatorial problem, it has recently been recognized that several sensor selection problems can be formulated as submodular optimization problems. In the context of estimating continuous random variables, it has been shown for linear measurement models that the negative of the log-determinant of the error covariance of the maximum a posteriori estimator (MAP) is submodular and thus amenable to optimization using greedy algorithms [11, 18–20]. This observation was recently extended to nonlinear measurement models by using the inverse of the Fisher Information Matrix as a proxy for the error covariance [14]. In the setting of sensor placement, it was proposed in [7] to select sensors to minimize the conditional entropy of the a posteriori distribution and in [15] to minimize the Kullback–Leibler divergence between the a-posteriori and a-priori distributions, both leading to submodular optimizations closely related to the one proposed here. However, we shall see that in the context of minimizing the probability of error for binary detection problems, the bound proposed here is typically tighter.

A stochastic submodular optimization was recently considered in [2, 13] for the maximization of functions of the form $f(X) \coloneqq \mathrm{E}[F(X, z)]$ where z is a random variable and $F(X, z)$ a function that is submodular in X, for every realization of z. More recently, [8, 17] considered the case where $f(X)$ is submodular, but not necessarily $F(X, z)$, which is the case we will encounter in our paper and was not allowed in the previous references. Both [17] and [8] use stochastic continuous greedy algorithms, rather than the discrete greedy algorithm that we propose here. We will show that the computational complexity of our algorithm compares favorably to these alternative algorithms when the number of sensors is large.

II. Problem Formulation

Consider the problem of estimating the value of a Bernoulli random variable $\theta \in \{0, 1\}$, while minimizing a criteria of the form

$$A\,\mathrm{P}(\hat{\theta} = 1, \theta = 0) + B\,\mathrm{P}(\hat{\theta} = 0, \theta = 1), \tag{1}$$

where $\hat{\theta}$ denotes the estimate and $A \geqslant 0$, $B \geqslant 0$ parameters that establish the cost of a false positive and a false detection, respectively. Without loss of generality, we assume that $A + B = 1$.

To estimate θ we have available L sensors whose measurements can be viewed as random variables

$$\{y_\ell : \ell \in \mathcal{L}\}, \quad \mathcal{L} \coloneqq \{1, 2, \ldots, L\}$$

that depend on θ. However, the estimator $\hat{\theta}$ can only use the measurements produced by a subset of the sensors

$$y_X \coloneqq \{y_\ell : \ell \in X\}, \quad X \subset \mathcal{L}.$$

Given a specific selection of sensors X, the construction of the optimal estimator based on y_X is straightforward as summarized in the following proposition that we prove in the technical report [10].

Proposition 1: For a given subset X of $\mathcal{L}$, the cost (1) is lower bounded by

$$\begin{aligned} J(X) &\coloneqq A\,\mathrm{P}(\hat{\theta} = 1, \theta = 0) + B\,\mathrm{P}(\hat{\theta} = 0, \theta = 1) \\ &\geqslant J^*(X) = \mathrm{E}\Big[\min\big\{Ap(0|y_X), Bp(1|y_X)\big\}\Big], \end{aligned} \tag{2}$$

with equality for the *optimal estimator*

$$\hat{\theta}(y_X) = \begin{cases} 1 & Ap(0|y_X) \leqslant Bp(1|y_X) \\ 0 & \text{otherwise,} \end{cases} \tag{3}$$

where $p(\theta|y_X)$ denotes the conditional distribution of θ given the measurements y_X produced by the sensors in X. □

The key problem addressed in this paper is the selection of a set $X \subset \mathcal{L}$ with no more than K elements that minimizes the cost $J^*(X)$ in (2) associated with the optimal estimator:

$$\min_{X \subset \mathcal{L} : |X| \leqslant K} J^*(X), \tag{4}$$

with the minimization carried out over all subsets of $\mathcal{L}$ with no more than K elements.

The minimization in (4) is combinatorial, which is complicated by the fact that the map $X \mapsto J^*(X)$ does not appear to have any of the properties that usually simplify this type of optimization. In fact, one can show by example that it is neither supermodular nor submodular, even when the measurements y_ℓ, $j \in \mathcal{L}$ are conditionally independent of each other given θ. However, we shall see shortly that $J^*(X)$ can be upper bounded in terms of a submodular function, which is easier to optimize.

Remark 1 (Sensor Reveal Games): The problem formulation above assumed that θ is a Bernoulli random variable with known a-priori distribution. In IUU this is often not the case, as the decision to engage in fishing is determined by another decision maker whose goals and capabilities may not be known. In addition, for AIS measurements the vessel's operator has the ability to manipulate the sensor, e.g., by turning it on and off. We have shown in [9] that the resulting detection problem can be viewed as a 2-player nonzero sum sensor-reveal game for which fictitious play is guaranteed to converge. In practice, this means that if the detector observes the actions of the other player, constructs an empirical average of the a-priori distribution of θ, and makes optimal decisions based on that estimated distribution (e.g., using the algorithms developed in this paper), the game is guaranteed to converge to a Nash equilibrium. □

III. Optimality Bound

To state the main result of this section, we recall that given two scalars $p_1, q_1 \in [0,1]$, the *Kullback–Leibler (KL) divergence* between the two Bernoulli distributions

$$p(\theta) = \begin{cases} p_1 & \theta = 1 \\ 1-p_1 & \theta = 0 \end{cases} \quad q(\theta) = \begin{cases} q_1 & \theta = 1 \\ 1-q_1 & \theta = 0 \end{cases}$$

is given by

$$\begin{aligned} D_{\mathrm{KL}}\big(p(\theta)\big\|q(\theta)\big) &= \delta_{\mathrm{KL}}(p_1, q_1) \\ &:= p_1 \log \frac{p_1}{q_1} + (1-p_1)\log\frac{1-p_1}{1-q_1}. \end{aligned} \quad (5)$$

The following result relates the optimal cost J^* in (2) with the KL-divergence between the a-posteriori distribution $p(\theta|y_X)$ of θ given y_X and an arbitrary Bernoulli distribution $\bar{p}(\cdot)$.

Theorem 1: Given a Bernoulli distribution $\bar{p}(\cdot)$, for every subset X of $\mathcal{L}$ we have that

$$\kappa J^*(X) \leqslant \eta - \mathrm{E}\Big[D_{\mathrm{KL}}\big(p(\theta|y_X)\big\|\bar{p}(\theta)\big)\Big], \quad (6)$$

with

$$\begin{aligned} \kappa &:= \frac{\eta - \delta_{\mathrm{KL}}\big(A, \bar{p}(1)\big)}{AB}, \\ \eta &:= \max\big\{-\log\bar{p}(0), -\log\bar{p}(1)\big\}. \end{aligned}$$

For the uniform distribution $\bar{p}(1) = 1/2$, we get

$$\kappa = \frac{1}{A}\log\Big(1+\frac{A}{B}\Big) + \frac{1}{B}\log\Big(1+\frac{B}{A}\Big), \quad \eta = \log 2.$$

□

In view of the bound (6), the optimal cost J^* can be decreased by selecting a subset $X \subset \mathcal{L}$ that maximizes

$$J_{\mathrm{sub}}(X) := \mathrm{E}\Big[D_{\mathrm{KL}}\big(p(\theta|y_X)\big\|\bar{p}(\theta)\big)\Big]. \quad (7)$$

The following result establishes that the map $X \mapsto J_{\mathrm{sub}}(X)$ is monotone and submodular, which will enable the use of tools from submodular maximization. We recall that a function $f : 2^{\mathcal{L}} \to \mathbb{R}$ that maps subsets of a finite set $\mathcal{L}$ to $\mathbb{R}$ is called *submodular* if for every $X \subset Y \subset \mathcal{L}$, $z \in \mathcal{L}\backslash Y$

$$f(X\cup\{z\}) - f(X) \geqslant f(Y\cup\{z\}) - f(Y) \quad (8)$$

and *monotone* if

$$X \subset Y \quad \Rightarrow \quad f(X) \leqslant f(Y).$$

Lemma 1: Assuming that all the random variables $\{y_\ell : \ell \in \mathcal{L}\}$ are conditionally independent given θ, for every $X \subset Y \subset \mathcal{L}$, $z \in \mathcal{L}\backslash Y$

$$J_{\mathrm{sub}}(X\cup\{z\}) - J_{\mathrm{sub}}(X) \geqslant 0 \quad (9)$$

$$J_{\mathrm{sub}}(X\cup\{z\}) - J_{\mathrm{sub}}(X) - J_{\mathrm{sub}}(Y\cup\{z\}) + J_{\mathrm{sub}}(Y) \geqslant 0, \quad (10)$$

which shows that the map $X \mapsto J_{\mathrm{sub}}(X)$ is monotone and submodular. Moreover,

$$\begin{aligned} J_{\mathrm{sub}}(X\cup\{z\}) - J_{\mathrm{sub}}(X) = \mathrm{E}\Big[\log\big(1+\rho(\theta|y_X)\big)\Big] \\ - \mathrm{E}\Big[\log\Big(1+\rho(\theta|y_X, y_z)\Big)\Big], \end{aligned} \quad (11)$$

with

$$\rho(\theta|y_Z) := \frac{1-p(\theta|y_Z)}{p(\theta|y_Z)}, \quad \forall Z \subset \mathcal{L},$$

where $\rho(\theta|y_Z)$ and the joint probabilities $p(y_Z, \theta)$ can be computed recursively using

$$p(y_X, y_z, \theta) = p(y_z\,|\,\theta)p(y_X, \theta), \quad (12)$$

$$\rho(\theta|y_X, y_z) = \frac{p(y_z|1-\theta)}{p(y_z|\theta)}\rho(\theta|y_X). \quad (13)$$

□

Remark 2: In the context of sensor placement, it was proposed in [15] to maximize the KL-divergence between the a-posteriori and a-priori distributions. For binary detection, this can be formally justified by Theorem 1, with the caveat that one generally gets a smaller upper bound in (6) by using the KL-divergence between the a-posterior and the *fixed uniform distribution* $\bar{p}(1) = 1/2$, rather than the a-priori distribution for θ. In essence, the goal here is not necessarily to select those measurements that lead to the largest distance from an a-priori to an a-posteriori distribution, but rather those measurements that minimize the total uncertainty within the a-posteriori distribution; and therefore we want to end up "far away" (in the KL-divergence sense) from the uniform distribution. □

The proofs of Theorem 1 and Lemma 1 can be found in the technical report [10].

IV. Stochastic Submodular Optimization

We are interested in optimizations of monotone submodular functions of the form:

$$\max\{f(X) : |X| \leqslant K, X \subset \mathcal{L}\} \quad (14)$$

The following algorithm can be used to approximately solve this optimization.

Algorithm 1 (Greedy with suboptimal selection): Given $\epsilon \geqslant 0$, $\gamma \geqslant 1$ the following algorithm returns a set X_K with K elements:

1) Initialize $X_0 = \varnothing$.
2) For each $k \in \{1, 2, \dots, K\}$:
   a) Pick an $x_k \in \mathcal{L}\backslash X_{k-1}$ for which

$$\begin{aligned} &\gamma\Big(\epsilon + f(X_{k-1}\cup\{x_k\}) - f(X_{k-1})\Big) \\ &\geqslant \max_{x\in\mathcal{L}\backslash X_{k-1}} f(X_{k-1}\cup\{x\}) - f(X_{k-1}) \end{aligned} \quad (15)$$

   b) Set $X_k = X_{k-1}\cup\{x_k\}$.

The following result adapted from [4] establishes a lower bound on the value of f on the final set X_K, as a function of the value of f on any other set X^*. The result is particularly useful for the set X^* that solves (14).

Theorem 2 (Deterministic): Assume that f is monotone, submodular, and $f(\varnothing) = 0$. For any set $X^* \subset \mathcal{L}$ with $|X^*| = K$ and every $\epsilon > 0, \gamma \geqslant 1$, the set X_K returned by Algorithm 1 satisfies

$$f(X_K) \geqslant \left(1 - \left(1 - \frac{1}{\gamma K}\right)^K\right)\big(f(X^*) - \gamma K \epsilon\big). \qquad \square$$

Consider now a monotone submodular function $f : 2^{\mathcal{L}} \to \mathbb{R}$ of the form

$$f(X) = \mathrm{E}[F(X, z)] \tag{16}$$

where $F : 2^{\mathcal{L}} \times \mathcal{Z} \to \mathbb{R}$ and z is a random variable taking values in $\mathcal{Z}$. The following greedy algorithm selects each point x_k based on an empirical average of $F(X, z)$.

Algorithm 2 (Greedy with stochastic optimal selection): Given $z_1, \dots, z_m$ independent identically distributed random variables with the same distribution as z, the following algorithm returns a set X_K with K elements:

1) Initialize $X_0 = \varnothing$.
2) For each $k \in \{1, 2, \dots, K\}$:
   a) Pick an $x_k \in \mathcal{L} \backslash X_{k-1}$ for which

   $$f_M(X_{k-1} \cup \{x_k\}) = \max_{x \in \mathcal{L} \backslash X_{k-1}} f_M(X_{k-1} \cup \{x\})$$

   where

   $$f_M(X) \coloneqq \frac{1}{M} \sum_{m=1}^{M} F(X, z_m) \tag{17}$$

   b) Set $X_k = X_{k-1} \cup \{x_k\}$. $\square$

Note that Algorithm 2 operates on a fixed set of random variables $z_1, \dots, z_m$ which retain their values across iterations of the algorithm. It is also important to note that, while we assume that the expected value in (16) is submodular, the empirical average in (17) may not be submodular.

Theorem 3 (Stochastic): Assume that f is monotone, submodular, and $f(\varnothing) = 0$. For any set $X^* \subset \mathcal{L}$ with $|X^*| = K$ and every $\bar{\epsilon} > 0$, the set X_K returned by Algorithm 2 satisfies

$$\mathrm{P}\left(f(X_K) \geqslant \left(1 - \left(1 - \frac{1}{K}\right)^K\right)\big(f(X^*) - \bar{\epsilon}\big)\right) \geqslant 1 - \delta \tag{18}$$

with

$$\delta \coloneqq K L\, e^{-\frac{1}{2} \frac{\bar{\epsilon}^2 M}{K^2 \sigma^2 + K \Delta \bar{\epsilon}/3}}, \tag{19}$$

where

$$\sigma^2 \coloneqq \max_{x_1, x_2, z} \mathrm{Var}\left[G_k(x_1, x_2, z)\right],$$

$$\Delta \coloneqq \max_{x_1, x_2, z} G_k(x_1, x_2, z) - \mathrm{E}\Big[G_k(x_1, x_2, z)\Big]$$

$$G_k(x_1, x_2, z) \coloneqq F(X_{k-1} \cup \{x_1\}, z) - F(X_{k-1} \cup \{x_2\}, z).$$

$\square$

In view of the relationship between M and δ in (19), it is straightforward to show that, for any given $\delta > 0$, it is always possible to guarantee that (18) holds by choosing

$$M \geqslant 2\Big(\frac{K^2 \sigma^2}{\bar{\epsilon}^2} + \frac{K \Delta/3}{\bar{\epsilon}}\Big) \log \frac{KL}{\delta}. \tag{20}$$

Remark 3 (Complexity): Algorithm 2 requires $O(MKL)$ evaluations of the function $F(\cdot)$ in (16). In view of (20), this leads to complexity

$$O\Big(\Big(\frac{K^3 L \sigma^2}{\bar{\epsilon}^2} + \frac{K^2 L \Delta/3}{\bar{\epsilon}}\Big) \log \frac{KL}{\delta}\Big).$$

The continuous SCG++ algorithms proposed in [8] could also be used to solve the stochastic submodular optimization that arises here. The complexity bound provided for their algorithm is $O(K^3 L^2 / \bar{\epsilon}^2)$, which could be significantly worst for large L. In the context of our detection problem, a large value for L corresponds to scenarios with a large number of sensors to choose from. $\square$

A. Proof of Theorem 3

The following lemma provides the main technical result needed to prove Theorem 3.

Lemma 2: For every $\epsilon > 0$,

$$\mathrm{P}\left(f(X_{k-1} \cup \{x_k\}) < \max_{x \in \mathcal{L} \backslash X_{k-1}} f(X_{k-1} \cup \{x\}) - \epsilon\right) \leqslant L e^{-\frac{1}{2} \frac{\epsilon^2 M}{\sigma^2 + \Delta \epsilon/3}} \tag{21}$$

$\square$

Proof of Lemma 2. Define

$$x^* \coloneqq \arg\max_{x \in \mathcal{L} \backslash X_{k-1}} f(X_{k-1} \cup \{x\}),$$

$$x^\dagger = \arg\max_{x \in \mathcal{L} \backslash X_{k-1}} f_M(X_{k-1} \cup \{x\}),$$

with the understanding that some tie-breaking algorithm is used to select the values of these variables when the maximum is not unique. In this case, $x^\dagger$ is a deterministic function of the random variables z_m and is therefore a random variable.

The probability in the left-hand side of (21) can be expressed in terms of $x^*, x^\dagger$ as

$$\mathrm{P}\left(x^\dagger \in \mathcal{L}_{\mathrm{bad}}[\epsilon]\right) = \sum_{x \in \mathcal{L}_{\mathrm{bad}}[\epsilon]} \mathrm{P}\left(x^\dagger = x\right), \tag{22}$$

where

$$\mathcal{L}_{\mathrm{bad}}[\epsilon] \coloneqq \big\{x \in \mathcal{L} \backslash X_{k-1} : f(X_{k-1} \cup \{x\}) - f(X_{k-1}) < f(X_{k-1} \cup \{x^*\}) - f(X_{k-1}) - \epsilon\big\}.$$

In view of (16), we can write

$$\mathcal{L}_{\mathrm{bad}}[\epsilon] = \big\{x \in \mathcal{L} \backslash X_{k-1} : \mathrm{E}\big[G_k(x, x^*, z)\big] < -\epsilon\big\}. \tag{23}$$

The definition of $x^\dagger$ guarantees that

$$\frac{1}{M} \sum_{m=1}^{M} F(X_{k-1} \cup \{x^\dagger\}, z_m) \eqqcolon f_M(X_{k-1} \cup \{x^\dagger\})$$

$$\geqslant f_M(X_{k-1} \cup \{x^*\}) \coloneqq \frac{1}{M} \sum_{m=1}^{M} F(X_{k-1} \cup \{x^*\}, z_m)$$

and therefore, for any $x \in \mathcal{L} \backslash X_{k-1}$, we have that

$$
\begin{aligned}
&\mathrm{P}(x^\dagger = x)\\
&\leqslant \mathrm{P}\left(\sum_{m=1}^{M} F(X_{k-1} \cup \{x\}, z_m) \geqslant \sum_{m=1}^{M} F(X_{k-1} \cup \{x^*\}, z_m)\right)\\
&\qquad = \mathrm{P}\left(G_k^M(x, x^*, z_1, \dots, z_m) \geqslant 0\right), \quad (24)
\end{aligned}
$$

where

$$
G_k^M(x, x^*, z_1, \dots, z_M) \coloneqq \frac{1}{M} \sum_{m=1}^{M} G_k(x, x^*, z_m).
$$

Further, since all the z_m are independent, we can use Chernoff's upper tail inequality in [5, Theorem 2.10] to conclude that, for every $\lambda \geqslant 0$,

$$
\mathrm{P}\left(G_k^M(x, x^*, z_1, \dots) - \mathrm{E}[G_k^M(x, x^*, z_1, \dots)] \geqslant \lambda\right) \leqslant e^{-\frac{1}{2}\frac{\lambda^2 M}{\sigma^2 + \lambda\Delta/3}}. \quad (25)
$$

Since all the z_m have the same distribution as z, we conclude from (23) that

$$
\mathrm{E}\left[G_k^M(x, x^*, z_1, \dots, z_M)\right] = \mathrm{E}\left[G_k(x, x^*, z)\right] < -\epsilon \leqslant 0,
$$

$\forall x \in \mathcal{L}_{\text{bad}}[\epsilon]$. Therefore, setting $\lambda = -\mathrm{E}\left[G_k(x, x^*, z)\right] \geqslant 0$ in (25) leads to

$$
\mathrm{P}\left(G_k^M(x, x^*, z_1, \dots, z_M) \geqslant 0\right) \leqslant e^{-\frac{1}{2}\frac{\mathrm{E}[G_k(x, x^*, z)]^2 M}{\sigma^2 + \Delta|\mathrm{E}[G_k(x, x^*, z)]|/3}}.
$$

Moreover, since the right-hand side is monotone decreasing with respect to the absolute value of $\mathrm{E}[G_k(x, x^*, z)]$ and this can never be smaller than ϵ, we conclude from this, (22), and (24) that

$$
\mathrm{P}\left(x^\dagger \in \mathcal{L}_{\text{bad}}[\epsilon, \gamma]\right) \leqslant \sum_{x \in \mathcal{L}_{\text{bad}}[\epsilon, \gamma]} e^{-\frac{1}{2}\frac{\gamma^2 \epsilon^2 M}{\sigma^2 + \Delta\gamma\epsilon/3}},
$$

from which (21) follows. ∎

Proof of Theorem 3. In view of Lemma 2, the probability that (15) will not hold with $\gamma = 1$ and $\epsilon = \bar{\epsilon}/K$ for at least one of the K iterations of Algorithms 2 is no larger than

$$
KL\, e^{-\frac{1}{2}\frac{\bar{\epsilon}^2 M}{K^2\sigma^2 + K\Delta\bar{\epsilon}/3}} = \delta.
$$

The result then follows from Theorem 2, applied to the submodular function $f(X)$ in (16), for which we have just established that (15) holds for all K iterations with probability, at least $1 - \delta$. ∎

V. Detection of IUU Fishing

In the context of IUU fishing detection, we are interested in estimating the value of a random variable θ_v that reflects whether or not a vessel v is engaging in illegal fishing. The measurements $\{y_\ell : \ell \in \mathcal{L}\}$ can be obtained by a variety of sensors, but we focus our attention on sensors that examine automatic identification system (AIS) tracks and satellite imagery. AIS tracks typically include the ship's identity and GPS-based position, compass heading, speed, and rates of turn. For both AIS and satellite imagery, different measurements correspond to different selections of a region of space and time over which the AIS or satellite data will be analyzed. AIS data is relatively cheap to access, but very prone to missed detections because vessels can turn off their AIS transponders [6, 9]. Satellite imagery is harder to manipulate but expensive to collect on a large scale. As selecting those measurements y_X that minimize the optimal probability of error in (2) is difficult, we minimize instead the upper bound in (6), or equivalently maximize the expected value of the KL-divergence in (7).

Our measurement model assumes that the overall spatial region of interest and interval of time have been partitioned into L spatio-temporal segments $\{s_\ell : \ell \in \mathcal{L}\}$, each corresponding to a small spatial cell and sub-interval of time. For a ship v and a spatio-temporal segment s_ℓ, we assume that the classification algorithms that run on AIS and/or satellite imagery collected on s_ℓ can return one of three values:

$$
y_\ell = \begin{cases} 0 & v \text{ not detected in } s_\ell, \\ 1 & v \text{ detected in } s_\ell, \text{ not fishing}, \\ 2 & v \text{ detected in } s_\ell, \text{ fishing}. \end{cases}
$$

However, the classification algorithms are not perfect and can produce misclassifications according to the following error probabilities:

$$
\begin{aligned}
p_{\text{loc}}^1 &\coloneqq \mathrm{P}(y_\ell = 1 \,|\, v \text{ not in } s_\ell)\\
p_{\text{loc}}^2 &\coloneqq \mathrm{P}(y_\ell = 2 \,|\, v \text{ not in } s_\ell)\\
p_{\text{loc}}^0 &\coloneqq \mathrm{P}(y_\ell = 0 \,|\, v \text{ in } s_\ell \text{ not fishing})\\
p_{\text{loc}}^0 &\coloneqq \mathrm{P}(y_\ell = 0 \,|\, v \text{ in } s_\ell \text{ and fishing})\\
p_{\text{cla}}^1 &\coloneqq \mathrm{P}(y_\ell = 1 \,|\, v \text{ in } s_\ell \text{ and fishing})\\
p_{\text{cla}}^2 &\coloneqq \mathrm{P}(y_\ell = 2 \,|\, v \text{ in } s_\ell \text{ not fishing}).
\end{aligned}
$$

In essence, $p_{\text{loc}}^0, p_{\text{loc}}^1, p_{\text{loc}}^2$ can be viewed a localization errors, and p_{cla}^1, p_{cla}^2 as classification errors. Assuming that the y_ℓ are conditionally independent of θ, given any of the events in the definition of the error probabilities above, we can use the law of total probability to compute the conditional distributions of an individual measurement y_ℓ given θ_v, which appear in the right-hand sides of (12)–(13):

$$
\begin{aligned}
p(0\,|\,0) &= (1 - p_{\text{loc}}^1 - p_{\text{loc}}^2)\big(1 - \mathrm{P}(v \text{ in } s_\ell \,|\, \theta_v = 0)\big)\\
&\quad + p_{\text{loc}}^0\, \mathrm{P}(v \text{ in } s_\ell \,|\, \theta_v = 0)\\
p(0\,|\,1) &= (1 - p_{\text{loc}}^1 - p_{\text{loc}}^2)(1 - \mathrm{P}(v \text{ in } s_\ell \,|\, \theta_v = 1))\\
&\quad + p_{\text{loc}}^0\, \mathrm{P}(v \text{ in } s_\ell \,|\, \theta_v = 1)\\
p(1\,|\,0) &= p_{\text{loc}}^1\big(1 - \mathrm{P}(v \text{ in } s_\ell \,|\, \theta_v = 0)\big)\\
&\quad + (1 - p_{\text{loc}}^0 - p_{\text{cla}}^2)\, \mathrm{P}(v \text{ in } s_\ell \,|\, \theta_v = 0)\\
p(1\,|\,1) &= p_{\text{loc}}^1(1 - \mathrm{P}(v \text{ in } s_\ell \,|\, \theta_v = 1))\\
&\quad + (1 - p_{\text{loc}}^0 - p_{\text{cla}}^2)\, \mathrm{P}(v \text{ in } s_\ell \text{ not fishing} \,|\, \theta_v = 1)\\
&\quad + p_{\text{cla}}^1\, \mathrm{P}(v \text{ in } s_\ell \text{ and fishing} \,|\, \theta_v = 1)\\
p(2\,|\,0) &= p_{\text{loc}}^2\big(1 - \mathrm{P}(v \text{ in } s_\ell \,|\, \theta_v = 0)\big)\\
&\quad + p_{\text{cla}}^2\, \mathrm{P}(v \text{ in } s_\ell \,|\, \theta_v = 0)\\
p(2\,|\,1) &= p_{\text{loc}}^2(1 - \mathrm{P}(v \text{ in } s_\ell \,|\, \theta_v = 1))\\
&\quad + p_{\text{cla}}^2\, \mathrm{P}(v \text{ in } s_\ell \text{ not fishing} \,|\, \theta_v = 1)
\end{aligned}
$$

$p^0_{\text{loc}} = 5\%$	$p^1_{\text{loc}} = 2\%$	$p^2_{\text{loc}} = 2\%$
$p^1_{\text{cla}} = 30\%$	$p^2_{\text{cla}} = 30\%$	$\eta = 100$ hours
$p_0 = 1$	$\lambda = .5$	$p_3 = 1$
$\tau_0 = .25$ deg	$\tau_1 = 5$ deg	$\tau_2 = 5$ deg

TABLE I: Parameters used for the numerical results

$$+ (1 - p^0_{\text{loc}} - p^1_{\text{cla}})\, \mathrm{P}(v \text{ in } s_\ell \text{ and fishing} \,|\, \theta_v = 1),$$

where

$$\begin{aligned}
&\mathrm{P}(v \text{ in } s_\ell \text{ not fishing} \,|\, \theta_v = 1) \\
&\quad = (1 - \mathrm{P}(\text{fishing} \,|\, \theta_v = 1, v \text{ in } s_\ell))\, \mathrm{P}(v \text{ in } s_\ell \,|\, \theta_v = 1) \\
&\mathrm{P}(v \text{ in } s_\ell \text{ and fishing} \,|\, \theta_v = 1) \\
&\quad = \mathrm{P}(\text{fishing} \,|\, \theta_v = 1, v \text{ in } s_\ell)\, \mathrm{P}(v \text{ in } s_\ell \,|\, \theta_v = 1).
\end{aligned}$$

To compute the conditional distribution above, we need three "behavioral parameters": (i) the probability that a vessel engaged in IUU fishing will be in s_ℓ, (ii) the probability that a vessel *not* engaged in IUU fishing will be in s_ℓ, and (iii) the probability that a vessel engaged in IUU fishing and in s_ℓ will actually fish. For illustration purposes, we assumed the following model

$$\begin{aligned}
&\mathrm{P}(v \text{ in } s_\ell \,|\, \theta_v = 0) = p_0 e^{-\text{distance from } s_\ell \text{ to declared path}/\tau_0} \\
&\mathrm{P}(v \text{ in } s_\ell \,|\, \theta_v = 1) = p_1 e^{-\text{distance from } s_\ell \text{ to declared path}/\tau_1} \\
&\quad p_1 \coloneqq \lambda + (1-\lambda) e^{-\text{distance from } s_\ell \text{ to good fishing}/\tau_2} \\
&\mathrm{P}(\text{fishing} \,|\, \theta_v = 1, v \text{ in } s_\ell) = \frac{p_3 \text{ historic fish. effort in } s_\ell/\eta}{1 + \text{historic fish. effort in } s_\ell/\eta},
\end{aligned}$$

which essentially assumes: (i) the probability that a ship deviates from its declared path decreases exponentially with the distance to the path, with the caveat that for vessels engaged in IUU fishing this probability also decreases with the distance to good fishing grounds; and (ii) the probability that an IUU fishing vessel will fish in s_ℓ increases monotonically with the historic record of fishing effort in the region s_ℓ, measured in number of hours of fishing per month during the period of interest.

Figure 1 shows the paths of two vessels and the corresponding sensor selections obtained from Algorithm 2 to maximize the expected value of the KL-divergence in (7). We recall that the expectation in (7) is taken with respect to the measurements y_X that would have been obtained using the sensors $X \subset \mathcal{L}$. Algorithm 2, replaces this expected value by an empirical average that we computed using $M = 3000$ Monte Carlo samples consistent with the measurement model described in the previous paragraph. For the historic fishing effort, we used data from Global Fishing Watch (GFW), for the month of June 2016 for which vessel tracks are available with position resolutions of 1/100th of a degree [1]. The regions s_ℓ used are approximately square corresponding to 1/5th of a degree in latitude and longitude. The remaining parameters used for these results can be found in Table I. The regions selected by Algorithm 2 reflect a combination of regions in the path of the vessels and nearby regions that would be attractive for a vessel engaged in IUU fishing.

VI. Conclusions

We addressed optimal sensor selection in estimating binary random variables. Our approach involves minimizing an upper bound on the error probability, which can be converted into the maximization of a submodular function. We then provide a randomized algorithm for which we provide sample complexity bounds. We presented numerical results related to the detection of IUU fishing using AIS tracks or satellite imagery. These results are based on real fishing effort data, but we have used synthetic parameters for the sensors. For AIS sensors, the estimation of these parameters should be possible based on data available from GWF.

References

[1] URL https://globalfishingwatch.org.

[2] A. Asadpour, H. Nazerzadeh, and A. Saberi. Stochastic submodular maximization. In *Int. Workshop on Internet and Network Economics*, pages 477–489. Springer, 2008.

[3] J. Baker. Transparent transhipping: detecting illegal fishing with satellite data. *Ship Technology*, 2018.

[4] L. Chamon and A. Ribeiro. Approximate supermodularity bounds for experimental design. In *Advances in Neural Information Processing Systems*, pages 5403–5412, 2017.

[5] F. Chung and L. Lu. Old and new concentration inequalities. *Complex Graphs and Networks*, 107:23–56, 2006.

[6] S. Emmert. Will fishing vessels find ways to avoid detection via AIS now that it's possible to track them so accurately? *Global Fishing Watch*, 2018.

[7] C. Guestrin, A. Krause, and A. P. Singh. Near-optimal sensor placements in Gaussian processes. In *Proc. of the 22nd Int. Conf. on Machine learning*, pages 265–272, 2005.

[8] H. Hassani, A. Karbasi, A. Mokhtari, and Z. Shen. Stochastic conditional gradient++. *arXiv preprint arXiv:1902.06992*, 2019.

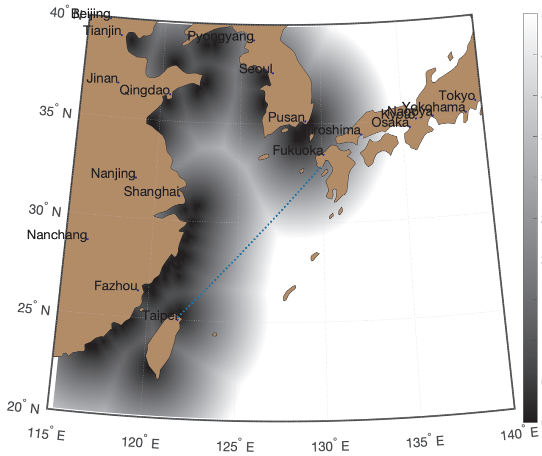
[9] J. P. Hespanha and D. Garagic. Sensor-reveal games. In *Proc. of the 58th Conf. on Decision and Contr.*, Dec. 2019.

[10] J. P. Hespanha and D. Garagić. Optimal sensor selection for binary detection based on stochastic submodular optimization: Extended version. Technical report, University of California, Santa Barbara, Santa Barbara, Mar. 2020.

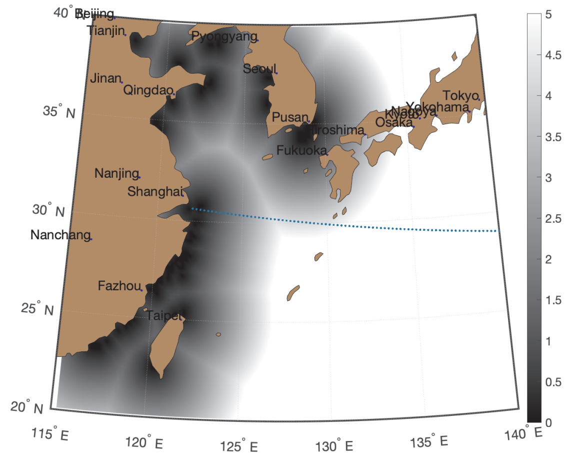
[11] S. T. Jawaid and S. L. Smith. Submodularity and greedy algorithms in sensor scheduling for linear dynamical systems. *Automatica*, 61:282–288, 2015.

[12] Y. Ji, S. Lee, E. Downing, W. Wang, M. Fazzini, T. Kim, A. Orso, and W. Lee. Rain: Refinable attack investigation with on-demand inter-process information flow tracking. In *Proc. of the 2017 ACM SIGSAC Conf. on Computer and Communications Security*, pages 377–390. ACM, 2017.

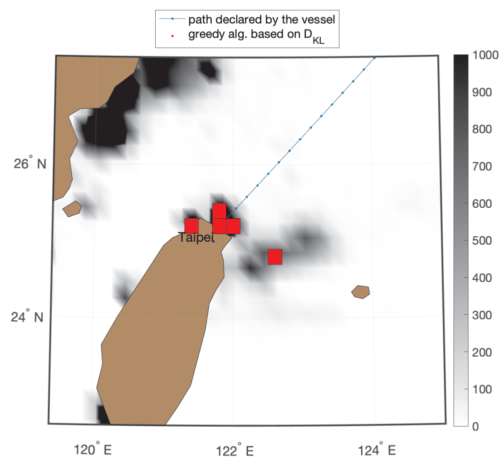
[13] M. Karimi, M. Lucic, H. Hassani, and A. Krause. Stochastic submodular maximization: The case of cov-



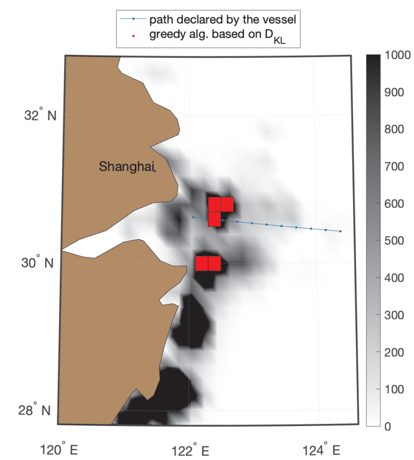
(a) Nagasaki, Japan to Keelung, Taiwan



(b) Shanghai, China to Lázaro Cárdenas, Mexico (partial view)



(c) regions selected for analysis in path from Nagasaki



(d) regions selected for analysis in path from Shanghai

Fig. 1: The top maps show synthetically generated paths for two vessels, one traveling from Nagasaki, Japan to Keelung, Taiwan (a) and another from Shanghai, China to Lázaro Cárdenas, Mexico (b). The bottom maps show the corresponding 5 regions selected by Algorithm 2 for sensor-data processing. The paths are represented by blue lines and the regions by red squares. The background colors in (a) and (b) represent distance to areas of large fishing effort and the background colors in (c) and (d) represents historic fishing effort in hours per month.

erage functions. In *Advances in Neural Information Processing Systems*, pages 6853–6863, 2017.

[14] M. Kirchner, D. Garagic, and J. P. Hespanha. Heterogeneous measurement selection for vehicle tracking using submodular optimization. In *Proc. of the 2020 IEEE Aerospace Conf.*, May 2020.

[15] A. Krause, A. Singh, and C. Guestrin. Near-optimal sensor placements in Gaussian processes: Theory, efficient algorithms and empirical studies. *J. of Machine Learning Research*, 9(Feb.):235–284, 2008.

[16] A. A. Kurekin, B. R. Loveday, O. Clements, G. D. Quartly, P. I. Miller, G. Wiafe, and K. Adu Agyekum. Operational monitoring of illegal fishing in Ghana through exploitation of satellite earth observation and AIS data. *Remote Sensing*, 11(3):293, 2019.

[17] A. Mokhtari, H. Hassani, and A. Karbasi. Conditional gradient method for stochastic submodular maximization: Closing the gap. In *Proc. of the 21st Int. Conf. on Artificial Intelligence and Statistics (AISTATS)*, volume 84, 2017.

[18] M. Shamaiah, S. Banerjee, and H. Vikalo. Greedy sensor selection: Leveraging submodularity. In *Proc. of the 49th Conf. on Decision and Contr.*, pages 2572–2577. IEEE, 2010.

[19] V. Tzoumas, A. Jadbabaie, and G. J. Pappas. Sensor placement for optimal Kalman filtering: Fundamental limits, submodularity, and algorithms. In *2016 American Control Conference (ACC)*, pages 191–196, 2016.

[20] H. Zhang, R. Ayoub, and S. Sundaram. Sensor selection for Kalman filtering of linear dynamical systems: Complexity, limitations and greedy algorithms. *Automatica*, 78:202–210, 2017.