Supplementary Materials for Bayesian Multiple Target Localization

To support the proof of Theorem 1 and 2, we need to following lemma, which provides an expression for the expected entropy after additional questions. First of all, we introduce some notations. For any pair of random variables W, V, we define $H(W \| V)$ to be the random variable taking the value

$$-\int_{-\infty}^{\infty} p(w|v) \log p(w|v)\, dw \tag{1}$$

for each $V = v$, assuming the conditional density function $p(w|v)$ exists. And $H(W|V)$ is the formal conditional entropy.

In addition, for any random variables W, V, U, we define $I(W; V \| U)$ to be the random variable taking the value

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} p(w, v|u) \log \frac{p(w, v|u)}{p(w|u)p(v|u)}\, dv\, dw, \tag{2}$$

for each $U = u$, assuming the conditional density functions exist. And $I(W; V|U)$ is the formal conditional mutual information.

Lemma 1. *Under any policy π, for all $n \geq 0$,*

$$E[H(p_{n+1})|X_{1:n}] = H(p_n) - I(\theta; X_{n+1} \| X_{1:n}). \tag{3}$$

Moreover,

$$E[H(p_N)] = H(p_0) - \sum_{n=0}^{N-1} I(\theta; X_{n+1}|X_{1:n}). \tag{4}$$

Proof. First of all, we prove the recursive relation (3). $H(p_n)$ is the entropy of the posterior distribution of θ, which is random through its dependence on the past history $X_{1:n}$, hence we can rewrite it as $H(p_n) = H(\theta \| X_{1:n})$. Similarly, $H(p_{n+1}) = H(\theta \| X_{1:n+1}) = H(\theta \| X_{1:n}, X_{n+1})$. Since all three terms in (3) are $\sigma(X_{1:n})$-measurable random variables, it suffices to prove (3) holds for any fixed history $X_{1:n} = x_{1:n}$, i.e.

$$E[H(\theta \| X_{1:n}, X_{n+1})|x_{1:n}] = H(\theta|x_{1:n}) - I(\theta; X_{n+1}|x_{1:n}). \tag{5}$$

Using information theoretic arguments, we have

$$\begin{aligned}
& E[H(\theta \| X_{1:n}, X_{n+1})|x_{1:n}] && \text{(6a)}\\
& = H(\theta|X_{n+1}, x_{1:n}) && \text{(6b)}\\
& = H(\theta, X_{n+1}|x_{1:n}) - H(X_{n+1}|x_{1:n}) && \text{(6c)}\\
& = H(\theta|x_{1:n}) + H(X_{n+1}|\theta, x_{1:n}) - H(X_{n+1}|x_{1:n}) && \text{(6d)}\\
& = H(\theta|x_{1:n}) - I(\theta; X_{n+1}|x_{1:n}) && \text{(6e)}
\end{aligned}$$

where (6b) comes from the definition of conditional entropy and (6c), (6d) come from the chain rule for conditional entropy. (6e) holds due to the relationship between entropy and mutual information. This proves (5).

Now, taking the expectation over $X_{1:n}$ on both sides of (3),

$$E\left[E[H(p_{n+1})|X_{1:n}]\right] = E[H(p_n)] - E\left[H(X_{n+1} \| X_{1:n})\right]. \tag{7}$$

Note that $E\left[E[H(p_{n+1})|X_{1:n}]\right] = E[H(p_{n+1})]$ by the iterated conditioning property of conditional expectation. Moreover, $E\left[I(\theta; X_{n+1} \| X_{1:n})\right] = I(\theta; X_{n+1}|X_{1:n})$ according to the definition of conditional entropy in (2). Hence, (7) is equivalent to

$$E[H(p_{n+1})] = E[H(p_n)] - I(\theta; X_{n+1}|X_{1:n}). \tag{8}$$

Applying (8) iteratively for $n = N-1, \ldots, 0$, we obtain (4), which concludes the proof. □

Proof of Theorem 1

Proof. According to Lemma 1, it suffices to prove that $I(\theta; X_{n+1}|X_{1:n}) \leq C_k \leq \log(k+1)$ for all $n \geq 0$ under any valid policy π. Since X_{n+1} depends on θ only through Z_{n+1}, we have

$$\begin{aligned}
I(\theta; X_{n+1}|X_{1:n}) &= I(Z_{n+1}; X_{n+1}|X_{1:n})\\
&= H(X_{n+1}|X_{1:n}) - H(X_{n+1}|Z_{n+1}, X_{1:n})\\
&= H(X_{n+1}|X_{1:n}) - H(X_{n+1}|Z_{n+1}).
\end{aligned} \tag{9}$$

Also, we have

$$H(X_{n+1}|X_{1:n}) = H\left(\sum_{z=0}^{k} \pi(z) f(\cdot|z)\right), \tag{10}$$

where $\pi(z)$ denotes the marginal distribution of Z_{n+1} and $f(\cdot|z)$ is the conditional probability density (mass) function of X_{n+1} given Z_{n+1}. Moreover,

$$H(X_{n+1}|Z_{n+1}) = \sum_{z=0}^{k} \pi(z) H\left(f(\cdot|z)\right). \tag{11}$$

Substituting (10) and (11) into (9) gives

$$\begin{aligned} & I(\theta; X_{n+1}|X_{1:n}) \\ &= H\left(\sum_{z=0}^{k} \pi(z) f(\cdot|z)\right) - \sum_{z=0}^{k} \pi(z) H\left(f(\cdot|z)\right) \\ &\leq \sup_q H\left(\sum_{z=0}^{k} q(z) f(\cdot|z)\right) - \sum_{z=0}^{k} q(z) H\left(f(\cdot|z)\right) \\ &= C_k, \end{aligned} \tag{12}$$

where $q(\cdot)$ is any probability mass function over $\{0, \dots, k\}$.

To see the second inequality, we note that the channel capacity C_k is bounded from above by the capacity of a noiseless channel, i.e.

$$C_k \leq I(Z_{n+1}, Z_{n+1}) = H(Z_{n+1}). \tag{13}$$

Since Z_{n+1} is a discrete random variable over $\{0, \dots, k\}$, the maximum possible value for the entropy $H(Z_{n+1})$ is obtained when X_{n+1} has a uniform distribution over $\{0, \dots, k\}$. Therefore, $H(Z_{n+1}) \leq \log(k+1)$, which completes the proof. □

Proof of Theorem 2

Proof. We first show that the noiseless answers $Z_{1:n}$ are iid under the dyadic policy. Let $U_{i,j}$ be iid $\text{Bernoulli}(1/2)$ random variables and let V_i be iid $\text{Uniform}(0, 2^{-N-1})$. Then $T_i := \sum_{j=1}^{N} 2^{-j} U_{i,j} + V_i$ are iid $\text{Uniform}(0,1)$. By the inversion method for simulation, $Q(T_i) = F_0^{-1}(T_i)$ provides a random variable that has cdf F_0, and so is equal in distribution to θ_i. Because T_i is independent across i, and θ_i is independent across i, the vector $(Q(T_i) : i = 1, \dots, k)$ is equal in distribution to θ. Each Z_n, considered as a function of θ, is equal in distribution to $U_{1,n} + ... + U_{k,n}$. Moreover, the vector $(Z_n : n = 1, \dots, N)$ is equal in distribution to the vector $(U_{1,n} + ... + U_{k,n} : n = 1, \dots, N)$, which is iid across n. Thus, $Z_{1:N}$ are iid.

Now, according to Lemma 1, it suffices to prove that under the dyadic policy, $I(\theta; X_{n+1}|X_{1:n}) = D_k$ for all $n \geq 0$. Under the dyadic policy, the noiseless answer $Z_{n+1} \sim \text{Bin}\left(k, \frac{1}{2}\right)$ and is independent of the previous history $X_{1:n}$ (this is a consequence of the independence of Z_{n+1} from $Z_{1:n}$ shown above). Hence, the marginal distribution function of Z_{n+1} is $\pi(z) = \binom{k}{z} \frac{1}{2^k}$. The remainder of the proof is similar to the proof of Theorem 1. □

To support the proof of Theorem 3, we introduce here some additional notation and derive an explicit formula for the posterior distribution after observing noiseless answers.

Consider a fixed n, where $1 \leq n \leq N$. For each binary sequence $s = \{s_1, \dots, s_n\}$, define

$$C_s = \left(\bigcap_{1 \leq j \leq n; s_j = 1} A_j\right) \bigcap \left(\bigcap_{1 \leq j \leq n; s_j = 0} A_j^c\right) \bigcap \text{supp}(f_0). \tag{14}$$

The collection $\{C_s : C_s \neq \emptyset, s \in \{0,1\}^n\}$ provides a partition of the support of f_0. A history of n questions provides information on which sets C_s contain which targets among $\theta_{1:k}$.

We will think of a sequence of binary sequences $s^{(1)}, \dots, s^{(k)}$ as a sequence of codewords indicating the sets in which each of the targets $\theta_{1:k}$ reside, i.e, indicating that θ_1 is in $C_{s^{(1)}}$, θ_2 is in $C_{s^{(2)}}$, etc. We may consider each binary sequence $s^{(1)}, \dots, s^{(k)}$ to be a column vector, and place them into an $n \times k$ binary matrix, $\mathcal{S}$. This binary matrix then codes the location of all k targets, and is a codeword for their joint location.

Moreover, to characterize the location of the random vector $\theta = (\theta_{1:k})$ in terms of its codeword $\mathcal{S}$, define $C_{\mathcal{S}} \subset \mathbb{R}^k$ to be the Cartesian product

$$C_{\mathcal{S}} = C_{s^{(1)}} \times \cdots \times C_{s^{(k)}}. \tag{15}$$

To be consistent with a noiseless answer Z_j, we must have exactly Z_j targets located in the question set A_j for each $1 \leq j \leq n$. This can be described in terms of a constraint on the matrix $\mathcal{S}$ as $s_j^{(1)} + \cdots + s_j^{(k)} = Z_j$, i.e., that the sum of the j^{th} row in the matrix $\mathcal{S}$ is Z_j. Thus, the set of all possible joint codewords that are consistent with $\{Z_{1:n} = z_{1:n}\}$ describing $\theta_{1:k}$ is

$$\begin{aligned} E_n = \{\mathcal{S} | s^{(1)}, \dots, s^{(k)} \in \{0,1\}^n, C_s^{(1)}, \dots, C_s^{(k)} \neq \emptyset, \\ s_j^{(1)} + \cdots + s_j^{(k)} = z_j, \text{for all } 1 \leq j \leq n\}. \end{aligned} \tag{16}$$

Now, we present the explicit characterization of the posterior distribution in the following lemma.

Lemma 2. *The posterior distribution given a sequence of noiseless answers $Z_{1:n} = z_{1:n}$ is*

$$p_n(u_{1:k}) = \frac{p_0(u_{1:k})}{p_0\left(\bigcup_{\mathcal{S} \in E_n} C_{\mathcal{S}}\right)}, \text{ for } u_{1:k} \in \bigcup_{\mathcal{S} \in E_n} C_{\mathcal{S}}, \tag{17}$$

where for a measurable set A, $p_0(A)$ denotes the integral $\int_A p_0(u_{1:k})\, du_{1:k}$. Moreover,

$$p_0\left(\bigcup_{\mathcal{S} \in E_n} C_{\mathcal{S}}\right) = \sum_{\mathcal{S} \in E_n} f_0(C_{s^{(1)}}) \dots f_0(C_{s^{(k)}}), \tag{18}$$

where $f_0(C_{s^{(i)}})$ *denotes the integral* $\int_{C_{s^{(i)}}} f_0(u)\,du$.

Proof of Theorem 3

Proof. First, we prove the result for noiseless answers. Under the dyadic policy, we partition $(0,1]$ into 2^N subintervals at time N. Now let's consider the event $\{\theta_i \in C|Z_{1:N}=z_{1:N}\}$, where C is one of such subintervals.

Let's denote the support of the posterior distribution $p_N(u_{1:k})$ by $D=\bigcup_{\mathcal{S}\in E_N} C_{\mathcal{S}}$. Moreover, denote the collection of matrices $\mathcal{S}\in E_N$ that are consistent with the event $\{\theta_i\in C|Z_{1:N}=z_{1:N}\}$ by $E_N(C)$. Note that $p_0(C_{\mathcal{S}}) = f_0(C_{s^{(1)}})f_0(C_{s^{(2)}})\dots f_0(C_{s^{(k)}}) = 2^{-Nk}$ under the dyadic policy. For simplicity, define $D_C = \bigcup_{\mathcal{S}\in E_N(C)} C_{\mathcal{S}}$. Therefore, using Lemma 2, we can compute the probability of $P(\theta_1\in C|Z_{1:N}=z_{1:N})$ as

$$\begin{aligned}
&P(\theta_1\in C|Z_{1:N}=z_{1:N})\\
&=\int_{u_{1:k}\in D_C} p_N(u_{1:k})\,du_{1:k}\\
&=\int_{u_{1:k}\in D_C}\frac{p_0(u_{1:k})}{\sum\limits_{\mathcal{S}\in E_N} f_0(C_{s^{(1)}})f_0(C_{s^{(2)}})\dots f_0(C_{s^{(k)}})}\,du_{1:k}\\
&=\sum_{\mathcal{S}\in E_N(C)}\frac{1}{2^{Nk}|E_N|}\int_{u_{1:k}\in C_{\mathcal{S}}} p_0(u_{1:k})\,du_{1:k}\\
&=\frac{|E_N(C)|}{|E_N|},
\end{aligned}\tag{19}$$

where $|E_N(C)|, |E_N|$ denote the cardinalities of $E_N(C), E_N$, respectively. Consider the construction of the $N\times k$ binary matrix $\mathcal{S}\in E_N$. The only requirement it needs to satisfy is that the sum of nth row is equal to z_n. Hence, we can construct the matrix row by row. Note that in step n, there are $\binom{k}{z_n}$ ways to specify the nonzero entries in the n^{th} row, for $n=1,2,\dots,N$. Thus, by the product rule,

$$|E_N| = \prod_{n=1}^{N}\binom{k}{z_n}. \tag{20}$$

Using combinatorial techniques, we have

$$|E_N(C)| = \prod_{n=1}^{N}\binom{k-1}{z_n-s_n}1_{\{0\le z_n-s_n\le k-1\}}. \tag{21}$$

Combining (19), (20) and (21) together and using the fact that $\theta_1,\theta_2,\dots,\theta_k$ are exchangeable,

$$\begin{aligned}
P(\theta_i\in C|Z_{1:N}=z_{1:N}) &= \frac{\prod_{n=1}^{N}\binom{k-1}{z_n-s_n}1_{\{0\le z_n-s_n\le k-1\}}}{\prod_{n=1}^{N}\binom{k}{z_n}}\\
&=\prod_{n=1}^{N}\begin{cases}\frac{z_n}{k}, & \text{if } s_n=1\\ 1-\frac{z_n}{k}, & \text{if } s_n=0\end{cases},
\end{aligned}\tag{22}$$

for $i=1,2,\dots,k$.

Equivalently,

$$P\left(\theta_i\in C|Z_{1:N}=z_{1:N}\right) = \prod_{n=1}^{N}\left(\frac{z_n}{k}\right)^{s_n}\left(1-\frac{z_n}{k}\right)^{1-s_n}. \tag{23}$$

Now we extend this result to the case with noisy answers. Firstly, we have

$$\begin{aligned}
P(\theta_i\in C|x_{1:N}) &= \sum_{z_{1:N}} P(\theta_i\in C, z_{1:N}|x_{1:N})\\
&=\sum_{z_{1:N}} P(\theta_i\in C|z_{1:N},x_{1:N})P(z_{1:N}|x_{1:N})\\
&=\sum_{z_{1:N}} P(\theta_i\in C|z_{1:N})P(z_{1:N}|x_{1:N}).
\end{aligned}\tag{24}$$

Under the dyadic policy, $Z_1,\dots,Z_N$ are conditionally independent given the noisy observations $x_1,\dots,x_N$. Thus, $P(z_{1:N}|x_{1:N}) = \prod_{n=1}^{N} P(z_n|x_{1:N})$. Moreover, due to the special structure of the dyadic policy, Z_n is independent of Z_j for all $j\neq n, j=1,\dots,N$, thus implying Z_n is independent of X_j for all $j\neq n, j=1,\dots,N$. Hence, $P(z_n|x_{1:N}) = P(z_n|x_n)$. Therefore, $P(z_{1:N}|x_{1:N}) = \prod_{n=1}^{N} P(z_n|x_n)$. According to (23), we have

$$\begin{aligned}
&P(\theta_i\in C|x_{1:N})\\
&=\sum_{z_{1:N}}\left(\prod_{n=1}^{N}(\frac{z_n}{k})^{s_n}(1-\frac{z_n}{k})^{1-s_n}\right)P(z_{1:N}|x_{1:N})\\
&=\sum_{z_{1:N}}\prod_{n=1}^{N}(\frac{z_n}{k})^{s_n}(1-\frac{z_n}{k})^{1-s_n}\prod_{n=1}^{N}P(z_n|x_n)\\
&=\sum_{z_{1:N}}\left(\prod_{n=1}^{N}(\frac{z_n}{k})^{s_n}(1-\frac{z_n}{k})^{1-s_n}P(z_n|x_n)\right)\\
&=\prod_{n=1}^{N}\left(\sum_{z_n=0}^{K}(\frac{z_n}{k})^{s_n}(1-\frac{z_n}{k})^{1-s_n}P(z_n|x_n)\right).
\end{aligned}\tag{25}$$

Furthermore, according to the definition of e_n, we have

$$\begin{aligned}
&\sum_{z_n=0}^{K}(\frac{z_n}{k})^{s_n}(1-\frac{z_n}{k})^{1-s_n}P(z_j|x_n) \\
&= \begin{cases} \sum_{z_n=0}^{K}\frac{z_n}{k}P(z_n|x_n) = \frac{e_n}{k}, & \text{if } s_n = 1, \\ \sum_{z_n=0}^{K}(1-\frac{z_n}{k})P(z_n|x_n) = 1-\frac{e_n}{k}, & \text{if } s_n = 0. \end{cases} \\
&= \left(\frac{e_n}{k}\right)^{s_n}\left(1-\frac{e_n}{k}\right)^{1-s_n}
\end{aligned} \tag{26}$$

Substituting (26) into (25) proves the first claim in Theorem 3.

Finally,

$$E[N(C)|x_{1:N}] = \sum_{i=1}^{k} P(\theta_i \in C|x_{1:N}) = kP(\theta_i \in C|x_{1:N}), \tag{27}$$

and we complete the proof. □

Now, we prove the claim made in Section 4 regarding the approximation ratio in the noiseless case.

Lemma 3. $H(\text{Bin}(k, \frac{1}{2}))/\log(k+1) \geq \frac{1}{2}$.

Proof. $H(\text{Bin}(k, \frac{1}{2})) = H(\sum_{i=1}^{k} B_i)$, where B_i are iid Bernoulli($\frac{1}{2}$). By Theorem 1 in (**?**), (but expressing entropy in base 2 instead of base e),

$$2^{2H(\text{Bin}(k,\frac{1}{2}))} \geq k2^{2H(B_1)} = 4k.$$

This implies that $H(\text{Bin}(k, \frac{1}{2})) \geq \frac{1}{2}\log(4k)$ and

$$\frac{H(\text{Bin}(k, \frac{1}{2}))}{\log(k+1)} \geq \frac{1}{2}\frac{\log(4k)}{\log(k+1)} \geq \frac{1}{2}.$$

□

Detailed implementation of the EP Algorithm

A detailed implementation of the Entropy Pursuit (EP) algorithm is presented below.

Algorithm 1 Implementation of EP

Obtain noisy observations $x_1, ..., x_N$.
Generate E_N.
Create matrix D with dimension $2^N \times |E_N|$. Each row in D corresponds to one pixel C, each element in the row represents the number of instances in C as per each $S \in E_N$.
for $S_i \in E_N$ **do**
  Update $Column\_i$ of D with the number of instances at each pixel C as per S_i;
end for
$m = 0, D^{(0)} = D, E_N^{(0)} = E_N$.
repeat
  for each pixel C **do**
    For each $u = 0, \ldots, k$, evaluate

$$\begin{aligned} &P(N(C) = u|x_{1:n}) \\ &= \frac{|\{S \in E_N^{(m)} : \text{codes for } C \text{ appear } u \text{ times in } S\}|}{|E_N^{(m)}|}; \end{aligned} \tag{28}$$

    Evaluate

$$\begin{aligned} &H(N(C)|x_{1:n}) \\ &= -\sum_{u=o}^{k} P(N(C) = u|x_{1:n}) \log(P(N(C) = u|x_{1:n})); \end{aligned} \tag{29}$$

  end for
  $C^* = \arg\max_C \; H(N(C)|x_{1:n})$;
  Query the oracle and obtain $Answer_{C^*}^{(m)} = Oracle(C^*)$;
  for $S_i \in E_N^{(m)}$ **do**
    if S_i is incompatible with $Answer_{C^*}^{(m)}$ **then**
      Remove S_i from $E_N^{(m)}$;
      Remove $Column_i$ from $D^{(m)}$;
    end if
  end for
  m=m+1;
until $H(O_{C^*}|x_{1:n}) = 0$
The unique columns of $D^{(m)}$ give the estimated instances joint location .