# Lower Bounds on the Minimax Risk for the Source Localization Problem

Praveen Venkatesh (vpraveen@cmu.edu) and Pulkit Grover (pulkit@cmu.edu)
Electrical & Computer Engineering, and the Center for the Neural Basis of Cognition, Carnegie Mellon University

***Abstract*—The "source localization" problem is one in which we estimate the location of a point source observed through a diffusive medium using an array of sensors. We obtain lower bounds on the minimax risk (mean squared-error in location) in estimating the location of the source, which apply to all estimators, for certain classes of diffusive media, when using a uniformly distributed sensor array. We show that for sensors of a fixed size, the lower bound decays to zero with increasing numbers of sensors. We also analyze a more physical sensor model to understand the effect of shrinking the size of sensors as their number increases to infinity, wherein the bound saturates for large sensor numbers. In this scenario, it is seen that there is greater benefit to increasing the number of sensors as the signal-to-noise ratio increases. Our bounds are the first to give a scaling for the minimax risk in terms of the number of sensors used.**

## I. Introduction

The source localization problem arises in many fields in different forms. Our principal motivation comes from Electroencephalography (EEG), a non-invasive brain measurement modality which uses electrodes placed on the scalp to sense electric potentials produced by neuronal activity within the brain [2]. The neurons (or groups of neurons) which produce this activity are usually modeled as current dipoles, and it is often of clinical or neuroscientific interest to estimate their locations [3]. For a single activated dipole within the brain (which is modeled as a point source), the electrodes sample a diffuse (blurred) representation of the dipole, as a result of the spatial low-pass-filtering effect of the different layers of the head, between the brain and scalp.

Our recent work [4] suggests that source localization can benefit greatly from an increase in sensor density. Revisiting the Nyquist rate arguments presented in [2] revealed that the question of how source imaging resolution is affected by sensor density has not been addressed. We argued intuitively that under appropriate conditions, the right algorithms might be able to recover the brain signal with much higher fidelity if more sensors were used. To follow up on these intuitive arguments, Grover provides an information-theoretic bound [5] on the accuracy of source localization algorithms. This information-theoretic bound, however, was derived for the unsampled, continuous-space potential on the scalp, or equivalently, in the limit of the number of sensors going to infinity. Hence, this bound does not substantiate the claim that source localization can benefit from increased sensor density.

Earlier work by Mosher et al. [6] also gave numerical lower bounds on source localization error, with *some* discussion on sensor density. However, the Cramer-Rao lower bounds derived by them do not apply to biased estimators, including several commonly used source localization algorithms [7], [8].

This paper seeks to address the shortcomings of earlier methods by deriving lower bounds on the minimax risk for the source localization problem, which apply to *all* estimators (biased *and* unbiased), and which capture how the bound scales with increasing numbers of sensors. Previous lower bounds are based on "spherical head models" [2], [4], which admit analytical solutions for the potential of a single dipole. But spherical models are not easily amenable to analytical lower bounds which seek to capture scaling in the number of sensors, because the spherical surface does not permit uniform sampling [9]. We therefore restrict our analysis to a source localization problem on a one-dimensional "circular" domain (to be described in detail in Section II). While this simplified model makes crucial assumptions (such as shift-invariance) that do not hold in realistic brain models, the techniques presented here extend naturally for deriving numerical bounds in more complex settings.

The one-dimensional toy problem is an effective tool for understanding bounds on source localization accuracy in other settings as well. A vast literature on linear inverse problems and deconvolution algorithms exists (see [10] for an introduction to this field), and has addressed the reduced one-dimensional problem in broad settings (see [11]–[13] and references therein). However, our setting and interpretation appears to be unique, since most prior work focuses on recovering a whole function (given certain smoothness constraints), rather than locating a point source [11]. Work that *does* address point sources and location-based error metrics [13, Ch. 7, Sec. 2] does not, to our knowledge, address scaling in the number of sensors.

Hence, we believe that this paper is the first to give lower bounds on the minimax error in estimating the location of a point source observed by sensors through a diffusive medium, and corrupted by additive white Gaussian noise. We are also the first to analyze the implications of physical sensor models

The full version of this paper is available online [1].

This work was supported in part by NSF CNS-1702694 and by Systems on Nanoscale Information fabriCs (SONIC), one of the six SRC STARnet Centers, sponsored by MARCO and DARPA. Praveen Venkatesh was supported (in part) by the Dowd Fellowship from the College of Engineering at Carnegie Mellon University. The authors would like to thank Philip and Marsha Dowd for their financial support and encouragement. The authors thank Prof. Anand Sarwate and Prof. Aarti Singh for their useful comments.

on such lower bounds. A simplistic theoretical model might assume that the number of sensors can increase without bound, while still maintaining a fixed SNR at each sensor. We show that this is non-physical in our setting, and derive numerical bounds which have radically different asymptotic behaviour.

We describe the problem setting in Section II and the minimax techniques and our main results in Section III. We conclude in Section IV with a discussion on the goodness of our bounds, and on possible extensions.

For EEG source localization, we intend to find algorithms that achieve scalings close to the bounds we present. We also expect to build upon the techniques presented in this paper to derive bounds for spherical and realistic brain models.

## II. Source Localization in 1D

### A. Description of the domain

We assume that the point source is located on a circle of circumference $S$. We use the variable '$s$' (for *s*pace) to represent a general point on this domain. We can also view this domain as a line, on which signals are periodic with period $S$. The point source is therefore located somewhere within one period. We denote the set of possible locations by $\Theta = [0, S)$.

### B. Sensor configuration

Sensors are assumed to be uniformly distributed over the domain, i.e., if there are $m$ sensors, they are placed at locations $s = 0,\ S/m,\ 2S/m,\ \ldots,\ (m-1)S/m$ (the offset of the first sensor is arbitrary, so without loss of generality we take it to be 0). The lower bounds we provide are for *this specific sensor configuration*. For a discussion on why this configuration might be an appropriate choice in the minimax setting, see Section IV.

### C. Signal model

All "continuous-space" signals (analogous to continuous- and discrete-*time* signals) on the aforementioned circular domain are of the form $f : \Theta \mapsto \mathbb{R}$. The point source located at $\theta$ is represented by the impulse signal, $f(s;\theta) = \delta(s-\theta)$, where $\delta(\cdot)$ is the Dirac delta function. The sensors observe this signal through the diffusive medium, which, intuitively speaking, blurs the impulse. More concretely, we assume that the medium is linear and shift-invariant, so that it can be characterized by a spatial impulse response. Then, blurring would correspond to low-pass filtering. Let this impulse response be given by $g(s)$. Then, the noiseless, continuous-space signal (post-filtering and pre-sampling) is given by the convolution, $x(s;\theta) = (g * f)(s) = g(s-\theta)$. Here, we further make the simplifying assumption that $g(s)$ has a sufficiently restricted support, so that aliasing effects are avoided, and $x(s;\theta)$ is always well-defined. To be precise, $g(s) = 0$ when $|s| > w/2$, where the "width" $w$ of the impulse response satisfies $w < S/2$ (the reason for the extra factor of $1/2$ will become clear in a later section). For the proof, $g(s)$ is also assumed to be Lipschitz continuous with parameter $\kappa$, i.e., $|g(u) - g(v)| \leq \kappa|u-v| \ \forall\ u, v \in \Theta$.

### D. Sensor and noise models

We consider two sensor models: a simplistic "point sensor" model and an "integrator sensor" model, to be used in Sections III-B and III-C respectively.

In the "point sensor" model, sensors *sample* the shifted continuous-space impulse response, with some additive noise. We denote the noiseless sampled version of $x(s;\theta)$ by the $m$-length vector $\underline{x}(\theta)$,

$$\underline{x}(\theta) = \left[x(0;\theta), \ldots, x\Big(\frac{kS}{m};\theta\Big), \ldots, x\Big(\frac{(m-1)S}{m};\theta\Big)\right]^T \quad (1)$$

where $m$ is the number of sensors and $k \in \{0, 1, \ldots, m-1\}$. Additive noise $\underline{\epsilon}$ corrupts $\underline{x}(\theta)$ to produce the noisy sensor values

$$\underline{y} = \underline{x}(\theta) + \underline{\epsilon}. \quad (2)$$

The noise $\underline{\epsilon}$ is zero-mean and Gaussian, $\underline{\epsilon} \sim \mathcal{N}(\underline{0}, \mathbf{\Sigma})$. We restrict our analysis to a "sensor noise" setting, wherein $\mathbf{\Sigma} = \sigma^2\mathbf{I}$, i.e., noise is i.i.d. across sensors.

In the "integrator sensor" model, sensors have a *width* which decreases as their number grows to infinity. Concretely, the $k$-th sensor which is located at $s = kS/m$ *occupies* the space $((k-\frac{1}{2})S/m, (k+\frac{1}{2})S/m)$. The signal sensed by this sensor is given by the $k$-th element of $\underline{x}(\theta)$,

$$x_k(\theta) = \int_{(k-\frac{1}{2})S/m}^{(k+\frac{1}{2})S/m} x(s;\theta)ds. \quad (3)$$

Here, a white noise process $\epsilon(s)$ additively corrupts the continuous-space signal $x(s;\theta)$ prior to sampling (see Fig. 1). $\underline{\epsilon}$ is generated by integrating $\epsilon(s)$ over each sensor's width. This leads to i.i.d. noise values at sensors (since sensors have no overlap), with variance inversely proportional to $m$, i.e., $\sigma(m) \propto 1/\sqrt{m}$ (refer Appendix A in the full version of this paper [1] for an explanation).

With the addition of Gaussian noise, each possible source location $\theta$ gives rise to a different distribution at the sensors, denoted by $P(\theta) = \mathcal{N}(\underline{x}(\theta), \mathbf{\Sigma})$. $\underline{y}$ is therefore one sample from the distribution $P(\theta)$. The space of distributions produced by all possible source locations is $\mathcal{P} = \{P(\theta) : \theta \in \Theta\}$.

In the source localization problem, we receive $n$ "trials"[1] of $\underline{y}$ according to equation (2), for the *same* point source $\theta$, using which we need to estimate the location of the source. Note that noise is assumed to be i.i.d. across trials, i.e., the $n$ realizations of $\underline{y}$ are independently drawn from $P(\theta)$, for the same location parameter $\theta$.

## III. Lower bounds on the minimax risk for the one-dimensional model

### A. Preliminaries

Several techniques for deriving lower bounds on the minimax risk are described in [14]. We follow the excellent tutorial of John Duchi [15] to outline the steps involved.

[1]The word "trials" comes from neuroscience literature, where multiple trials of an experiment are run, with the same stimulus being given to the participant in each trial, evoking the same response in the brain (with some trial-to-trial variability, which we ignore). For our purposes, we assume that trials are i.i.d. realizations of sensor noise.

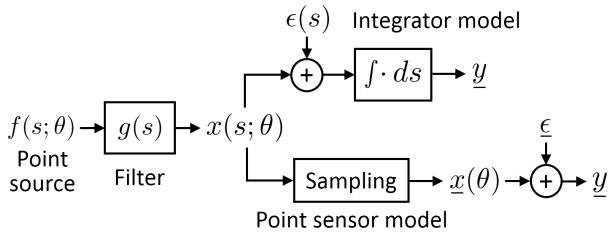


Fig. 1. A block diagram of the two sensor models described in section II-C.

Consider the following estimation problem: we have $n$ i.i.d. random samples $Y^n$ from a distribution $P$, which is indexed by a parameter $\theta \in \Theta$. Denote the set of these distributions by $\mathcal{P} = \{P(\theta) : \theta \in \Theta\}$. Suppose we now wish to estimate $\theta$ from $Y^n$. Define a loss metric $\rho(\theta, \hat{\theta})$ (e.g. $\rho(\theta, \hat{\theta}) = |\theta - \hat{\theta}|$). For this metric, we can define the minimax risk over all possible estimators $\hat{\theta}(Y^n)$ and all parameters $\theta \in \Theta$:

$$\mathfrak{M}_n(\mathcal{P}, \Phi \circ \rho) = \inf_{\hat{\theta}} \sup_{\theta \in \Theta} \mathbb{E}[\Phi \circ \rho(\hat{\theta}(Y^n), \theta)] \tag{4}$$

where $\Phi$ is any non-decreasing function (e.g. $\Phi(\rho) = \rho^2$).

We start by lower bounding the maximum risk of the estimation problem with the average risk of a multiple hypothesis testing problem. For this, we first need to define a $2\delta$-packing.

*Definition 1:* A set $\Theta_{\mathcal{V}} = \{\theta_v : v \in \mathcal{V}\}$ for some finite index set $\mathcal{V} \subset \mathbb{N}$ is said to be a $2\delta$-packing in the $\rho$-metric if

$$\rho(\theta_i, \theta_j) \geq 2\delta \qquad \forall\ \theta_i, \theta_j \in \Theta_{\mathcal{V}}. \tag{5}$$

*Theorem 1:* If we can find a $2\delta$-packing $\Theta_{\mathcal{V}}$ of $\Theta$, then we can lower bound the minimax estimation risk by the average testing risk:

$$\mathfrak{M}_n(\mathcal{P}, \Phi \circ \rho) \geq \Phi(\delta) \inf_{\psi} \mathbb{P}(\psi(Y^n) \neq V) \tag{6}$$

where $V$ is the unknown true hypothesis which takes values uniformly from $\mathcal{V}$, and $\psi$ is our estimate of the hypothesis.

For a proof of this theorem, we refer the reader to [14, Sec. 2.2], or to [15, Prop. 13.3]. Intuitively, Theorem 1 says that the error in estimating $\theta_i$ is likely to be large if it is difficult to distinguish $\theta_i$ from $\theta_j$, i.e. if the probability of error is high.

We now need to lower bound the probability of error in the hypothesis testing problem. The simplest way to do this is to consider a binary hypothesis test and use what is known as Le Cam's method:

*Theorem 2:* For a binary hypothesis test, i.e., $\mathcal{V} = \{0, 1\}$,

$$\inf_{\psi} \mathbb{P}(\psi(Y^n) \neq V) = 1 - \|P_0 - P_1\|_{TV} \tag{7}$$

where $P_i$ is short-hand for $P(\theta_i)$ and $\|P_0 - P_1\|_{TV}$ is the total variation distance between the two distributions, defined as $\|\cdot\|_{TV} = \frac{1}{2}\|\cdot\|_1$.

For a proof, we refer the reader to [14, Thm. 2.2], or to [15, Prop. 2.11] for a more readable account. Intuitively, Theorem 2 states that the minimum achievable error probability in a binary hypothesis testing problem is related to the distance between the distributions corresponding to the two hypotheses.

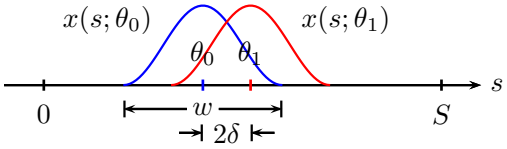


Fig. 2. Depiction of the continuous-space signals produced by two hypotheses, $\theta_0$ and $\theta_1$. Note that each signal is a shifted impulse response, and therefore has a support of size $w$. The set $\{\theta_0, \theta_1\}$ is a $2\delta$-packing of $\Theta$, so the signals are separated by a distance $2\delta$.

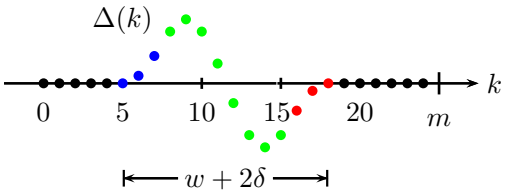


Fig. 3. Plot of the samples of the difference signal, $\Delta(k)$. The total size of the support of the difference signal is $w+2\delta$, hence at most $\left\lfloor \frac{(w+2\delta)m}{S} + 1 \right\rfloor$ samples of $\Delta(k)$ are non-zero.

Thus, the final lower bound can be written as

$$\mathfrak{M}_n(\mathcal{P}, \Phi \circ \rho) \geq \frac{\Phi(\delta)}{2}\left(1 - \|P_1^n - P_0^n\|_{TV}\right). \tag{8}$$

The superscripts "$n$" remind us that these are $n$-fold product distributions, since we have $n$ i.i.d. trials used in the estimate. The difficulty in Le Cam's method lies in selecting the two hypotheses to trade-off the effect due to a small value of $\delta$ and a large value of $\|P_1^n - P_0^n\|_{TV}$ appropriately, to derive the tightest possible bound.

### B. Lower bounds using Le Cam's method

For the source localization problem, we measure *risk* in terms of the squared-distance between the true location and the estimated location. Hence, our loss function is $\Phi(\rho(\theta, \hat{\theta})) = |\theta - \hat{\theta}|^2$, where $\hat{\theta}$ is some estimate of the location based on $n$ trials of the noisy sensor observations $\underline{y}$. We now state the main result of this paper.

*Theorem 3:* For a source localization problem as defined in Section II under the point sensor model, the minimax risk in estimating the location of a point source is lower bounded by

$$\mathfrak{M}_n(\mathcal{P}, \Phi \circ \rho) \geq \sup_{0 < \delta < \frac{S}{4}} \frac{\delta^2}{2}\left[1 - \sqrt{\frac{2n\kappa^2\delta^2}{\sigma^2}\left(\frac{(w+2\delta)m}{S} + 1\right)}\right].$$

For sufficiently large $m$, the bound can be approximated by

$$\mathfrak{M}_n(\mathcal{P}, \Phi \circ \rho) \gtrsim \frac{1}{32}\frac{\sigma^2 S}{nm\kappa^2 w}. \tag{9}$$

*Proof:* Starting from equation (8), we proceed to derive the total variation distance for the distributions of interest. Using Pinsker's inequality [14, Lemma 2.5] and the convenient tensorization of the KL-divergence [15], we see that

$$\|P_1^n - P_0^n\|_{TV}^2 \leq \frac{1}{2} D_{KL}(P_0^n \| P_1^n) = \frac{n}{2} D_{KL}(P_0 \| P_1). \tag{10}$$

For multivariate normal distributions with the same covariance, the KL-divergence is given by

$$D_{KL}(P_0 \| P_1) = (\underline{\mu}_0 - \underline{\mu}_1)^T \mathbf{\Sigma}^{-1} (\underline{\mu}_0 - \underline{\mu}_1), \tag{11}$$

where $\underline{\mu}_0$ and $\underline{\mu}_1$ are the means of $P_0$ and $P_1$ respectively [16]. For the case of sensor noise, $P_i = \mathcal{N}(\underline{x}(\theta_i), \sigma^2\mathbf{I})$, as described in Section II-D. Hence, combining equations (8), (10) and (11), we see that

$$\mathfrak{M}_n(\mathcal{P}, \Phi\circ\rho) \geq \frac{\delta^2}{2}\left[1 - \sqrt{\frac{n}{2\sigma^2}\|\underline{x}(\theta_0) - \underline{x}(\theta_1)\|^2}\right]. \quad (12)$$

Let $\underline{\Delta} \stackrel{\text{def.}}{=} \underline{x}(\theta_0) - \underline{x}(\theta_1)$ for brevity. Also, let $\Delta(k)$ denote the $k$-th element of $\underline{\Delta}$, and $\mathbb{I}_A(k)$ denote the indicator function of $k$ belonging to the set $A$ (i.e., $\mathbb{I}_A(k) = 1$ if $k \in A$, and 0 otherwise). Then, we have

$$\|\underline{\Delta}\|^2 = \sum_{k=0}^{m-1}|\Delta(k)|^2 = \sum_{k=0}^{m-1}|\Delta(k)|^2\,\mathbb{I}_{\{\ell:|\Delta(\ell)|\neq 0\}}(k) \quad (13)$$

$$= \sum_{k=0}^{m-1}\left|x\Big(\frac{kS}{m};\theta_0\Big) - x\Big(\frac{kS}{m};\theta_1\Big)\right|^2\mathbb{I}_{\{\ell:|\Delta(\ell)|\neq 0\}}(k) \quad (14)$$

$$= \sum_{k=0}^{m-1}\left|g\Big(\frac{kS}{m} - \theta_0\Big) - g\Big(\frac{kS}{m} - \theta_1\Big)\right|^2\mathbb{I}_{\{\ell:|\Delta(\ell)|\neq 0\}}(k) \quad (15)$$

where $x(s;\theta_i)$ is the continuous-space filtered signal described in Section II-C. For an impulse response $g$ which is Lipschitz continuous with parameter $\kappa$, we can upper bound the term within the summation:

$$\left|g\Big(\frac{kS}{m} - \theta_0\Big) - g\Big(\frac{kS}{m} - \theta_1\Big)\right| \leq \kappa|\theta_0 - \theta_1| = \kappa\cdot 2\delta \quad (16)$$

since $|\theta_0 - \theta_1| = 2\delta$ by virtue of the $2\delta$ packing. Hence,

$$\|\underline{\Delta}\|^2 \leq 4\kappa^2\delta^2\sum_{k=0}^{m-1}\mathbb{I}_{\{\ell:|\Delta(\ell)|\neq 0\}}(k) = 4\kappa^2\delta^2\|\underline{\Delta}\|_0. \quad (17)$$

$\|\underline{\Delta}\|_0$ is the number of non-zero elements in $\underline{\Delta}$, which is equal to the number of sensors in the total region covered by the signals $x(s;\theta_0)$ and $x(s;\theta_1)$ (see Figs. 2 and 3). Therefore, $\|\underline{\Delta}\|_0 = \lfloor\frac{(w+2\delta)m}{S} + 1\rfloor$, since at most a fraction $(w+2\delta)/S$ of the $m$ sensors (plus 1, to account for edge-effects) can lie in the region covered by the two impulse responses. The final upper bound on $\|\underline{\Delta}\|^2$ is hence

$$\|\underline{\Delta}\|^2 \leq 4\kappa^2\delta^2\left\lfloor\frac{(w+2\delta)m}{S} + 1\right\rfloor \leq 4\kappa^2\delta^2\left(\frac{(w+2\delta)m}{S} + 1\right). \quad (18)$$

Since sensors are uniformly distributed, and since the domain is periodic, this holds even if $\theta_0$ lies at the edge of the domain (close to $s = 0$, for example). In such a case, one part of the signal $x(s;\theta_0)$ will appear at the left edge of the domain, and the remaining part will appear as the repetition from the period $[S, 2S)$, at the right edge of the domain. Also note that the two signals $x(s;\theta_0)$ and $x(s;\theta_1)$ overlap at most once, since $w < S/2$, as stated in Section II-C.

Combining equations (12) and (18),

$$\mathfrak{M}_n(\mathcal{P}, \Phi\circ\rho) \geq \frac{\delta^2}{2}\left[1 - \sqrt{\frac{2n\kappa^2\delta^2}{\sigma^2}\left(\frac{(w+2\delta)m}{S} + 1\right)}\right]. \quad (19)$$

This bound holds for any feasible $2\delta$-packing. We can derive the tightest possible bound by maximizing the above quantity over all feasible $\delta$. Since the maximum attainable separation between two hypotheses on the periodic domain is $S/2$, we maximize $\delta$ over $(0, S/4)$. This yields the theorem.

To derive the approximation, we observe that for sufficiently large numbers of sensors, the optimal value of $\delta$ starts becoming small, as hypotheses must chosen to be closer in order to make the error probability large. Hence, neglecting edge effects and terms of order $\delta^3$, we can give an approximate bound for large $m$ using a heuristic: tighten the bound by choosing $\sqrt{\frac{2n\kappa^2\delta^2 wm}{\sigma^2 S}} = \frac{1}{2}$. This is achieved for $\delta = \sqrt{\frac{\sigma^2 S}{8nm\kappa^2 w}}$, so that for $\Phi(\delta) = \delta^2$, equation (19) becomes

$$\mathfrak{M}_n(\mathcal{P}, \Phi\circ\rho) \gtrsim \frac{1}{32}\frac{\sigma^2 S}{nm\kappa^2 w}. \quad (20)$$

This completes the proof. ■

### C. Towards a more physical sensor model

The bound in Theorem 3 may mislead one to think that the error in source localization goes to zero, as the number of sensors goes to infinity. This observation contradicts the information-theoretic lower bound given in [5], where even with an infinite number of sensors, the lower bound is strictly greater than zero. This indicates that the bound in Section III-B is failing to capture some effect which leads to saturation of error as the number of sensors increases.

Indeed, the bound in Theorem 3 assumes that an arbitrarily large number of sensors can be squeezed into a finite amount of space, and that these sensors *still* have the same SNR. This is non-physical, since in reality, each sensor has a certain width, and the value measured by it is the integral of the continuous-space signal within that width (analogous to a sample-and-hold circuit, which integrates a continuous-*time* signal over a short period). If the width of a sensor is decreased, then effectively, its SNR reduces (this is explained in Appendix A in the full version of this paper [1]).

This observation motivates the "integrator sensor" model defined in Section II-D. With this model, the bounding technique used in Section III-B no longer easily admits a closed-form solution which gives a bound for all values of $m$. We instead compute the bound numerically, using exact expressions for error probability in a binary hypothesis test. For Gaussian distributions of equal variance and a uniform prior on the hypotheses, the minimum possible error probability (over all tests) is a $Q$-function in the distance between their means [17, Ch. 7],

$$P_e = \inf_{\psi}\mathbb{P}(\psi(Y^n) \neq V) = Q\left(\frac{\|\underline{\mu}_0 - \underline{\mu}_1\|}{2\sigma(m)}\right), \quad (21)$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}}e^{-u^2/2}du$. Hence, from Theorem 1,

$$\mathfrak{M}_n(\mathcal{P}, \Phi\circ\rho) \geq \Phi(\delta)P_e = \delta^2 Q\left(\frac{\|\underline{x}(\theta_0) - \underline{x}(\theta_1)\|}{2\sigma(m)}\right). \quad (22)$$

This bound is valid for any $\{\theta_0, \theta_1\}$ constituting a $2\delta$-packing of $\Theta$. To numerically optimize over all possible packings, we relax the process by fixing $\theta_0 = 0$ and optimizing over $\theta_1 = 2\delta$. We compute sensor values for a triangular impulse

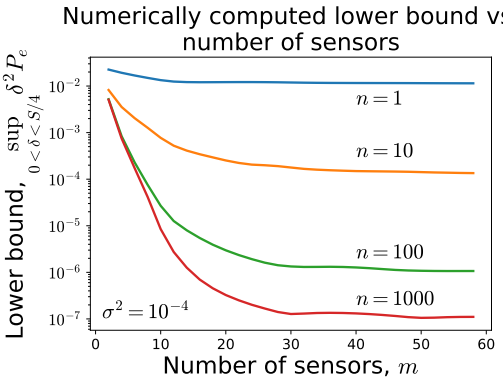


Fig. 4. A numerically generated plot of the lower bound on minimax risk of source localization, against the number of uniformly distributed sensors, $m$, following the "integrator sensor" model (code available online [1]). The different curves are for different numbers of "trials", $n$ (number of i.i.d. sensor observations). The bound saturates for large $m$, indicating that for a *fixed* number of trials, there is decreasing benefit to using more sensors. However, the plot suggests that with more trials, the bound saturates at increasingly larger numbers of sensors. Thus, there may be benefit to increasing the number of sensors when you have more trials.

response by performing numerical integration, and find the best possible $\delta$ by grid search.

Plotting this numerically-computed bound for different numbers of trials reveals some interesting trends (see Fig. 4). First, the bound saturates as $m$ grows large, matching the behaviour of the information-theoretic bound. Second, the lower bound on error is larger for smaller numbers of sensors, as we expect. Third and most importantly, as the number of trials increases, the bound begins to saturate only at a larger number of sensors. In other words, there is *increased* benefit in using more sensors, when we *also* have more trials. This substantiates the intuition given in [4]: a larger number of sensors will sample a larger number of high-frequency components of the source signal (and hence allow us to locate it better), but only if the noise floor is sufficiently low. Noise variance, in turn, can be reduced by averaging over multiple trials.

## IV. Discussion

While the lower bound presented in Section III-B is non-physical for a large number of sensors, it still gives some important insights. The bound tells us how the error scales with the number of sensors up to $m = \lfloor \frac{S}{c} \rfloor$, for sensors of a constant width $c$. Beyond this, averaging noise over a larger number of trials $n$ may help one achieve a target mean squared error in location. An analytical derivation of the trade-off between $m$ and $n$ in achieving a certain target MSE is relegated to future work. Further, more sophisticated techniques for deriving lower bounds, such as Fano's method [14, Sec. 2.7.1] are yet to be analyzed.

We also note that the minimax bounds from compressive sensing literature [18] *do not* apply in our setting, because those consider the problem of recovering the *whole* source signal, $f(s;\theta)$. The loss function used there is hence the energy of the difference signal, $\int |f(s) - \hat{f}(s)|^2 ds$. We, on the other hand, are interested in bounding the minimax error in *location*, i.e., $|\theta - \hat{\theta}|^2$.

While the bounds in Section III-B were derived for a one-dimensional circular domain, extensions to a 1D linear domain (of fixed length $S$) and to a $d$-dimensional domain (circular or linear) are fairly straightforward, under certain conditions. For a 1D linear domain, we require that sensors capture the entirety of the impulse response even at the edges of the domain. Hence, sensors must be placed uniformly in $[-w/2, S+w/2]$. For extending to $d$ dimensions, it suffices to apply Assouad's Lemma [14, Sec. 2.7.2] to separate the problem into $d$ independent 1D problems. Such extensions substantially broaden the application domains where this bound could be useful.

Also note, the bounds derived in this paper apply only when sensors are uniformly distributed. Intuitively, this is the *correct* choice in the minimax setting, since any non-uniform placement will have one or more points which is far from all sensors. The "maximization" over $\theta \in \Theta$ would then select a $\theta$ for which the value of the impulse response at the sensors is smallest, so that SNR is minimized.

## References


[1] Full version of this paper. [Online]. Available: `https://github.com/praveenv253/isit-2017`.

[2] P. Nunez and R. Srinivasan, *Electric Fields of the Brain: The Neurophysics of EEG*. Oxford University Press, 2006.

[3] S. Baillet, J. C. Mosher, and R. M. Leahy, "Electromagnetic brain mapping," *IEEE Signal Process. Mag.*, vol. 18, no. 6, pp. 14–30, 2001.

[4] P. Grover and P. Venkatesh, "An information-theoretic view of EEG sensing," *Proc. IEEE*, vol. 105, no. 2, pp. 367–384, February 2017.

[5] P. Grover, "Fundamental limits on source-localization accuracy of EEG-based neural sensing," in *ISIT*, July 2016, pp. 1794–1798.

[6] J. C. Mosher, M. E. Spencer, R. M. Leahy, and P. S. Lewis, "Error bounds for EEG and MEG dipole source localization," *EEG and Clin. Neurophysiol.*, vol. 86, no. 5, pp. 303 – 321, 1993.

[7] M. S. Hämäläinen and R. J. Ilmoniemi, "Interpreting magnetic fields of the brain: minimum norm estimates," *Medical & biological engineering & computing*, vol. 32, no. 1, pp. 35–42, 1994.

[8] F.-H. Lin, T. Witzel, S. P. Ahlfors, S. M. Stufflebeam, J. W. Belliveau, and M. S. Hämäläinen, "Assessing and improving the spatial accuracy in MEG source localization by depth-weighted minimum-norm estimates," *NeuroImage*, vol. 31, no. 1, pp. 160 – 171, 2006.

[9] T. L. Heath, *The Thirteen Books of Euclid's Elements*. Dover Publications, New York, 1956, vol. 3.

[10] G. Bal, *Introduction to inverse problems*, 2012, [Online]. Available: `http://www.columbia.edu/~gb2030/PAPERS/IntroductionInverseProblems.pdf`.

[11] L. Cavalier and A. Tsybakov, "Sharp adaptation for inverse problems with random noise," *Probab. Theory Relat. Fields*, pp. 323–354, 2002.

[12] S. Efromovich, "Robust and efficient recovery of a signal passed through a filter and then contaminated by non-Gaussian noise," *IEEE Trans. Inf. Theory*, vol. 43, no. 4, pp. 1184–1191, 1997.

[13] I. A. Ibragimov and R. Z. Has'minskii, *Statistical Estimation: Asymptotic Theory*, 1st ed. Springer-Verlag New York, 1981, vol. 16.

[14] A. B. Tsybakov, *Introduction to Nonparametric Estimation*. Springer New York, 2009.

[15] J. C. Duchi, *Information Theory and Statistics*. Stanford University, 2015, lecture Notes for Statistics 311/Electrical Engineering 377.

[16] J. Duchi, "Derivations for linear algebra and optimization," [Online]. Available: `https://web.stanford.edu/~jduchi/projects/general_notes.pdf`.

[17] L. L. Scharf, *Statistical Signal Processing*. Addison-Wesley, 1991.

[18] E. Arias-Castro, E. J. Candes, and M. A. Davenport, "On the fundamental limits of adaptive sensing," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 472–481, January 2013.

## APPENDIX A
## SNR IN THE INTEGRATOR MODEL

In the integrator model of sensors defined in Section II-D, recall that the samples are given by equation (3). Consider a sensor located at the point $s_0 \in [0, S)$ and with a width of $S/m$. That is, we assume each of the $m$ sensors occupies the maximum area that it can, $1/m$ of the total. As $m$ grows large, the size of each sensor must shrink in order to accommodate their larger number.

In order to remain consistent with our previous notation, we first define the *index*, $k$, of the sensor located at $s_0$. (Note that as $m$ increases, the index $k$ must grow while keeping the position of the sensor fixed, so that we're seeing how the SNR changes at the same point).

$$k = \Big\lfloor \frac{ms_0}{S} \Big\rceil, \tag{23}$$

where $\lfloor z \rceil$ is read as "round of $z$", denoting the closest integer to the point $z \in \mathbb{R}$. Note that as $m$ grows large, $kS/m \to s_0$.

We now compute the approximate SNR of the sensor located at $s_0$ as follows. The power of the signal is given by the squared absolute value of the samples,

$$|x_k(\theta)|^2 = \left( \int_{(k-\frac{1}{2})S/m}^{(k+\frac{1}{2})S/m} x(s;\theta) ds \right)^2 \tag{24}$$

$$\overset{(a)}{\approx} \left( \frac{S}{m}\, x\Big(\frac{kS}{m};\theta\Big) \right)^2 \tag{25}$$

$$\overset{(b)}{\approx} \frac{S^2}{m^2} \big(x(s_0;\theta)\big)^2, \tag{26}$$

where in approximation (a), we have assumed that the signal is roughly constant in the small interval $\big((k-\frac{1}{2})S/m, (k+\frac{1}{2})S/m\big)$, taking the value of the signal at $s = kS/m$. In approximation (b), we make use of the fact that for sufficiently large $m$, $kS/m \approx s_0$. Note a key step in the power computation: the integrand in equation (24) is *not* $|x(s;\theta)|^2$, because the sensor will average out the signal as a part of its sensing process.

Hence, irrespective of the location of the sensor, $s_0$, the signal power diminishes as $1/m^2$ since the sensor size *must* diminish at least as fast as $1/m$, if the number of sensors has to increase to infinity.

The noise variance of the sensor located at $s_0$ is computed similarly:

$$\mathbb{E}[\epsilon_k^2] = \mathbb{E}\left[ \left( \int_{(k-\frac{1}{2})S/m}^{(k+\frac{1}{2})S/m} \epsilon(s) \right)^2 \right] \tag{27}$$

$$= \mathbb{E}\left[ \int_{(k-\frac{1}{2})S/m}^{(k+\frac{1}{2})S/m} \epsilon(s) ds \int_{(k-\frac{1}{2})S/m}^{(k+\frac{1}{2})S/m} \epsilon(s') ds' \right] \tag{28}$$

$$= \mathbb{E}\left[ \int_{(k-\frac{1}{2})S/m}^{(k+\frac{1}{2})S/m} \int_{(k-\frac{1}{2})S/m}^{(k+\frac{1}{2})S/m} \epsilon(s)\epsilon(s') ds ds' \right] \tag{29}$$

$$= \int_{(k-\frac{1}{2})S/m}^{(k+\frac{1}{2})S/m} \int_{(k-\frac{1}{2})S/m}^{(k+\frac{1}{2})S/m} \mathbb{E}[\epsilon(s)\epsilon(s')] ds ds' \tag{30}$$

$$= \int_{(k-\frac{1}{2})S/m}^{(k+\frac{1}{2})S/m} \int_{(k-\frac{1}{2})S/m}^{(k+\frac{1}{2})S/m} R_\epsilon(s,s') ds ds' \tag{31}$$

$$= \int_{(k-\frac{1}{2})S/m}^{(k+\frac{1}{2})S/m} \int_{(k-\frac{1}{2})S/m}^{(k+\frac{1}{2})S/m} \sigma^2 \delta(s-s') ds ds' \tag{32}$$

$$= \int_{(k-\frac{1}{2})S/m}^{(k+\frac{1}{2})S/m} \sigma^2 ds \tag{33}$$

$$= \sigma^2 \frac{S}{m}, \tag{34}$$

where $R_\epsilon(s,s') = \sigma^2 \delta(s-s')$ is the autocorrelation function of the white noise process, and $\sigma^2$ is the amplitude of the white noise power spectral density.

We can understand how the SNR of the sensor at $s_0$ scales with $m$, therefore, by taking the ratio of (26) and (34)

$$\text{SNR}(s_0) = \frac{S^2}{m^2} \big(x(s_0;\theta)\big)^2 \cdot \frac{m}{S\sigma^2} \tag{35}$$

$$= \frac{S}{m\sigma^2} \big(x(s_0;\theta)\big)^2 \propto \frac{1}{m}. \tag{36}$$

Hence, as the number of sensors $m$ grows to infinity, their sizes must necessarily reduce at least as fast as $1/m$, and as a result, their SNR also falls at least as fast as $1/m$, when we view these sensors as integrators.