

Article

# Two New Families of Local Asymptotically Minimax Lower Bounds in Parameter Estimation

Neri Merhav 

The Viterbi Faculty of Electrical and Computer Engineering, Technion—Israel Institute of Technology, Technion City, Haifa 3200003, Israel; merhav@ee.technion.ac.il; Tel.: +972-4-8294737

**Abstract:** We propose two families of asymptotically local minimax lower bounds on parameter estimation performance. The first family of bounds applies to any convex, symmetric loss function that depends solely on the difference between the estimate and the true underlying parameter value (i.e., the estimation error), whereas the second is more specifically oriented to the moments of the estimation error. The proposed bounds are relatively easy to calculate numerically (in the sense that their optimization is over relatively few auxiliary parameters), yet they turn out to be tighter (sometimes significantly so) than previously reported bounds that are associated with similar calculation efforts, across many application examples. In addition to their relative simplicity, they also have the following advantages: (i) Essentially no regularity conditions are required regarding the parametric family of distributions. (ii) The bounds are local (in a sense to be specified). (iii) The bounds provide the correct order of decay as functions of the number of observations, at least in all the examples examined. (iv) At least the first family of bounds extends straightforwardly to vector parameters.

**Keywords:** parameter estimation; minimax estimation; mean-square error; hypothesis testing; probability of error; Fisher information



**Citation:** Merhav, N. Two New Families of Local Asymptotically Minimax Lower Bounds in Parameter Estimation. *Entropy* **2024**, *26*, 944. <https://doi.org/10.3390/e26110944>

Academic Editors: Raúl Alcaraz, Luca Faes, Leandro Pardo and Boris Ryabko

Received: 19 September 2024  
Revised: 29 October 2024  
Accepted: 4 November 2024  
Published: 4 November 2024



**Copyright:** © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The theory of parameter estimation consists of a very large plethora of lower bounds (as well as upper bounds), which characterize the fundamental performance limits of any estimator in a given parametric model. In this context, it is common to distinguish between Bayesian bounds (see, e.g., the Bayesian Cramér–Rao bound [1], the Bayesian Bhattacharyya bound, the Bobrovsky–Zakai bound [2], the Bellini–Tartara bound [3], the Chazan–Zakai–Ziv bound [4], the Weiss–Weinstein bound [5,6], and more (see [7] for a comprehensive overview), and non-Bayesian bounds, where in the former, the parameter to be estimated is considered a random variable with a given probability law, as opposed to the latter, where it is assumed an unknown deterministic constant. The category of non-Bayesian bounds is further subdivided into two subclasses; one is associated with local bounds that hold for classes of estimators with certain limitations, such as unbiased estimators (see, e.g., the Cramér–Rao bound [8–12], the Bhattacharyya bound [13], the Barankin bound [14], the Chapman–Robbins bound [15], the Fraser–Guttman bound [16], the Keifer bound [17], and more), and the other is the subclass of minimax bounds (see, e.g., Ziv and Zakai [18], Hajek [19], Le Cam [20], Assouad [21], Fano [22], Lehmann (Sections 4.2–4.4 in [23]), Nazin [24], Yang and Barron [25], Guntuboyina [26,27], Kim [28], and many more).

In this paper, we focus on the minimax approach, and more concretely, on the local minimax approach. According to the minimax approach, we are given a parametric family of probability density functions (or probability mass functions, in the discrete case),  $\{p(x_1, \dots, x_n | \theta), (x_1, \dots, x_n) \in \mathbb{R}^n, \theta \in \Theta\}$ , where  $\theta$  is a  $d$ -dimensional parameter vector,  $\Theta \subseteq \mathbb{R}^d$  is the parameter space,  $n$  is a positive integer designating the number of

observations, and we define a loss function,  $\ell(\theta, \hat{\theta}_n)$ , where  $\hat{\theta}_n$  is an estimator, which is a function of the observations  $x_1, \dots, x_n$  only. The minimax performance is defined as

$$R_n(\Theta) \triangleq \inf_{\hat{\theta}_n(\cdot)} \sup_{\theta \in \Theta} \mathbb{E}_\theta \{ \ell(\theta, \hat{\theta}_n) \}, \quad (1)$$

where  $\mathbb{E}_\theta$  denotes expectation w.r.t.  $p(\cdot|\theta)$ . As customary, we consider here loss functions with the property that  $\ell(\theta, \hat{\theta}_n)$  depends on  $\theta$  and  $\hat{\theta}_n$  only via their difference, that is,  $\ell(\theta, \hat{\theta}_n) = \rho(\theta - \hat{\theta}_n)$ , where the function  $\rho(\cdot)$  satisfies certain assumptions (see Section 2). The local asymptotic minimax performance at the point  $\theta \in \Theta$  is defined as follows (see also, e.g., [19]). Let  $\{\zeta_n^*, n \geq 1\}$  be a positive sequence, tending to infinity, with the property that

$$r(\theta) \triangleq \lim_{\delta \downarrow 0} \liminf_{n \rightarrow \infty} \inf_{\hat{\theta}_n(\cdot)} \sup_{\{\theta' : \|\theta' - \theta\| \leq \delta\}} \zeta_n^* \cdot \mathbb{E}_{\theta'} \{ \rho(\hat{\theta}_n - \theta') \} \quad (2)$$

is a strictly positive finite constant. Then, we say that  $r(\theta)$  is the local asymptotic minimax performance with respect to (w.r.t.)  $\{\zeta_n^*\}$  at the point  $\theta \in \Theta$ . Roughly speaking, the significance is that the performance of a good estimator,  $\hat{\theta}_n$ , at  $\theta$  is about  $\mathbb{E}_\theta \{ \rho(\theta - \hat{\theta}_n) \} \approx r(\theta) / \zeta_n^*$ . For example, in the scalar mean square error (MSE) case, where  $\rho(\varepsilon) = \varepsilon^2$ , and where the observations are Gaussian, i.i.d., with mean  $\theta$  and known variance  $\sigma^2$ , it is actually shown in Example 2.4, p. 257 in [23] that  $r(\theta) = \sigma^2$  w.r.t.  $\zeta_n^* = n$ , for all  $\theta \in \mathbb{R}$ , which is attained by the sample mean estimator,  $\hat{\theta}_n = \frac{1}{n} \sum_{i=1}^n x_i$ .

Our focus in this work is on the derivation of some new lower bounds that are as follows: (i) essentially free of regularity conditions on the smoothness of the parametric family  $\{p(\cdot|\theta), \theta \in \Theta\}$ , (ii) relatively simple and easy to calculate, at least numerically, which amounts to the property that the bound contains only a small number of auxiliary parameters to be numerically optimized (typically, no more than two or three parameters), (iii) tighter than earlier reported bounds that are associated with similar calculation efforts as described in (ii), and (iv) lend themselves to extensions that yield even stronger bounds (albeit with more auxiliary parameters to be optimized), as well as extensions to vector parameters. We propose two families of lower bounds on  $R_n(\Theta)$ , along with their local versions, of bounding  $r(\theta)$ , with the four above-described properties. The first applies to any convex, symmetric loss function  $\rho(\cdot)$ , whereas the second is more specifically oriented to the moments of the estimation error,  $\rho(\varepsilon) = |\varepsilon|^t$ , where  $t$  is a positive real, not necessarily an integer, with special attention devoted to the MSE case,  $t = 2$ . For the sake of simplicity and clarity of the exposition, in the first two main sections of the paper (Sections 3 and 4), our focus is on the case of a scalar parameter, as most of our examples are associated with the scalar case. In Section 5, we extend some of our findings to the vector case.

To put this work in the perspective of earlier work on minimax estimation, we next briefly review some of the basic approaches in this problem area. Admittedly, considering the vast amount of literature on the subject, our review below is by no means exhaustive. For a more comprehensive review, the reader is referred to Kim [28].

First, observe the simple fact that the minimax performance is lower bounded by the Bayesian performance of the same loss function (see, e.g., [1–7]) for any prior on the parameter,  $\theta$ , and so, every lower bound on the Bayesian performance is automatically a valid lower bound also on the minimax performance (see Section 4.2 in [23]). Indeed, in Section 2.3 in [26], it is argued that the vast majority of existing minimax lower bounding techniques are based upon bounding the Bayes risk from below w.r.t. some prior. Many of these Bayesian bounds, however, are subjected to certain restrictions and regularity conditions concerning the smoothness of the prior and the family of densities,  $\{p(\cdot|\theta), \theta \in \Theta\}$ .

Dating back to Ziv and Zakai's 1969 article [18] on parameter estimation, applied mostly in the context of time-delay estimation, this prior puts all its mass equally on two values,  $\theta_0$  and  $\theta_1$ , of the parameter  $\theta$ , and considering an auxiliary hypothesis testing problem of distinguishing between the two hypotheses,  $\mathcal{H}_0 : \theta = \theta_0$  and  $\mathcal{H}_1 : \theta = \theta_1$  with equal priors (see Section 2 for exact definitions). A simple argument regarding the sub-optimality

of a decision rule that is based on estimating  $\theta$  and deciding on the hypothesis with the closer value of  $\theta$ , combined with Chebychev's inequality, yields a simple lower bound on the corresponding Bayes risk, and hence also the minimax risk, in terms of the probability of error of the optimal decision rule. Five years later, Bellini and Tartara [3], and then independently, Chazan, Zakai, and Ziv [4], improved the bound of [18] using somewhat different arguments and obtained Bayesian bounds that apply to the uniform prior. These bounds are also given in terms of the error probability pertaining to the optimal maximum a posteriori (MAP) decision rule of binary hypothesis testing with equal priors, but this time, it had an integral form. These bounds were demonstrated to be fairly tight in several application examples, but they are rather difficult to calculate in most cases. Shortly before the Bellini–Tartara and the Chazan–Zakai–Ziv articles were published, Le Cam [20] proposed a minimax lower bound, which is also given in terms of the error probability associated with binary hypothesis testing, or equivalently, the total variation between  $p(\cdot|\theta_0)$  and  $p(\cdot|\theta_1)$ , under the postulate that the loss function  $\ell(\cdot, \cdot)$  is a metric. We will refer to Le Cam's bound in a more detailed manner later, in the context of our first proposed bound.

A decade later, Assouad [21] extended Le Cam's two-point testing bound to multiple points, where instead of just two test points,  $\theta_0$  and  $\theta_1$ , as before, there are, more generally,  $m$  test points,  $\theta_0, \theta_1, \dots, \theta_{m-1}$ , and correspondingly, the auxiliary hypothesis testing problem consists of  $m$  hypotheses,  $\mathcal{H}_i : \theta = \theta_i, i = 0, 1, \dots, m - 1$ , with certain priors (again, to be defined in Section 2). Based on those test points, Assouad devised the so-called hypercube method. Another related bounding technique that is based on multiple test points, and referred to as Fano's method, amounts to further bounding from below the error probability of multiple hypotheses using Fano's inequality (see Section 2.10 in [29]). Considering the large number of auxiliary parameters to be optimized when multiple hypotheses are present, these bounds demand heavy computational efforts. Also, Fano's inequality is often loose, even though it is adequate enough for the purpose of proving converse-to-coding theorems in information theory [29]. In later years, Le Cam [30] and Yu [31] extended Le Cam's original approach to apply to testing mixtures of densities. More recently, Yang and Barron [25] related the minimax problem to the metric entropy of the parametric family,  $\{p(\cdot|\theta), \theta \in \Theta\}$ , and Cai and Zhou [32] combined Le Cam's and Assouad's methods by considering a larger number of dimensions. Guntuboyina [26,27] pursued a different direction by deriving minimax lower bounds using  $f$ -divergences.

The outline of this article is as follows. In Section 2, we define the problem setting, provide a few formal definitions along with background, and establish the notation. In Section 3, we develop the first family of bounds, and in Section 4, we present the second family, both for the scalar parameter case. Finally, in Section 5, we extend some of our findings to the vector case.

## 2. Problem Setting, Background, Definitions, and Notation

As outlined in the last paragraph of the Introduction, Sections 3 and 4 are on the scalar parameter case, whereas in Section 5, we extend some of the results to the case of a vector parameter with dimension  $d$ . In order to avoid repetition, we formalize the problem and define the notation here for the more general vector case, with the understanding that for the scalar case, all definitions remain the same, except that they are confined to the special case of  $d = 1$ .

Consider a family of probability density functions (PDFs),  $\{p(\cdot|\theta), \theta \in \Theta\}$ , where  $\theta$  is a parameter vector of dimension  $d$  to be estimated, and  $\Theta \subseteq \mathbb{R}^d$  is the parameter space. We denote by  $\mathbb{E}_\theta\{\cdot\}$  the expectation operator w.r.t.  $p(\cdot|\theta)$ . Let  $\mathbf{X} = (X_1, \dots, X_n)$  be a random vector of observations, governed by  $p(\cdot|\theta)$  for some  $\theta \in \Theta$ . The support of  $p(\cdot|\theta)$  is assumed  $\mathcal{X}^n$ , the  $n$ th Cartesian power of the alphabet  $\mathcal{X}$  of each component  $X_i, i = 1, \dots, n$ . The alphabet  $\mathcal{X}$  may be a finite set, a countable set, a finite interval, an infinite interval, or the entire real line. In the first two cases, the PDFs should be understood to be replaced by probability mass functions and integrations over the observation space should be replaced by summations. A realization of  $\mathbf{X}$  will be denoted by  $\mathbf{x} = (x_1, \dots, x_n) \in \mathcal{X}^n$ .

An estimator,  $\hat{\theta}_n$ , is given by any function of the observation vector,  $g_n : \mathcal{X}^n \rightarrow \Theta$ , that is,  $\hat{\theta}_n = g_n[x]$ . Since  $\mathbf{X}$  is random, so is the estimate  $g_n[\mathbf{X}]$ , as well as the estimation error,  $\varepsilon_n[\mathbf{X}] = \hat{\theta}_n - \theta = g_n[\mathbf{X}] - \theta$ . We associate with every possible vector value,  $\varepsilon$ , of  $\varepsilon_n[\mathbf{X}]$  a certain loss (or “cost”, or “price”),  $\rho(\varepsilon)$ , where  $\rho(\cdot)$  is a non-negative function with the following properties: (i) monotonically non-increasing in each component,  $\varepsilon_i$ , of  $\varepsilon$ , wherever  $\varepsilon_i < 0, i = 1, 2, \dots, d$ , (ii) monotonically non-decreasing in each component,  $\varepsilon_i$ , of  $\varepsilon$ , wherever  $\varepsilon_i \geq 0, i = 1, 2, \dots, d$ , and (iii)  $\rho(0) = 0$ .

Referring to Sections 3 and 4, which deals with the scalar case ( $d = 1$ ), we further adopt the following assumptions regarding the loss function  $\rho$ . In Section 3, we assume that  $\rho(\cdot)$  is as follows: (iv) convex, and (v) symmetric, i.e.,  $\rho(-\varepsilon) = \rho(\varepsilon)$  for every  $\varepsilon$ . In Section 4, we assume, more specifically, that  $\rho(\varepsilon) = |\varepsilon|^t$ , where  $t$  is a positive constant, not necessarily an integer. This is a special case of the class of loss functions considered in Section 3, except when  $t \in (0, 1)$ , in which case,  $|\varepsilon|^t$  is a concave (rather than a convex) function of  $\varepsilon$ . In Section 5, for the vector extension of the results of Section 3, the above-mentioned symmetry assumption (v) is extended to become radial symmetry; namely,  $\rho(\varepsilon)$  will be assumed to depend on  $\varepsilon$  only via its Euclidean norm,  $\|\varepsilon\|$ .

The expected cost of an estimator  $g_n$  at a point  $\theta \in \Theta$ , is defined as

$$R_n(\theta, g_n) \triangleq \mathbb{E}_\theta\{\rho(g_n[\mathbf{X}] - \theta)\}. \tag{3}$$

The global minimax performance is defined as

$$R_n(\Theta) \triangleq \inf_{g_n} \sup_{\theta \in \Theta} R_n(\theta, g_n). \tag{4}$$

Another related notion is that of local asymptotic minimax performance, defined in Section 1, and repeated here for the sake of completeness. Let  $\{\zeta_n^*, n \geq 1\}$  be a positive sequence, tending to infinity, with the property that  $r(\theta)$ , defined as in (2), is a strictly positive finite constant. Then, we say that  $r(\theta)$  is the local asymptotic minimax performance w.r.t.  $\{\zeta_n^*\}$  at  $\theta \in \Theta$ . The sequence  $\{1/\zeta_n^*\}$  is referred to as the convergence rate of the minimax estimator.

Similarly, as in earlier articles on minimax lower bounds, many of our proposed bounds are given in terms of the error probability pertaining to certain auxiliary hypothesis testing problems that are associated with two or more *test points* in the parameter space, and the choice of those test points is subjected to optimization. We therefore provide a few definitions, notation conventions, and background associated with elementary hypothesis testing. For further details, the reader is referred to any one of many textbooks that cover the topic, for example, Van Trees (see Sections 2.2 and 2.3 in [1]), Helstrom (see Chapter III in [33]), or Whalen (see Chapter 5 in [34]).

For given  $\theta_0$  and  $\theta_1$ , both in  $\Theta$ , consider the problem of deciding between two possible hypotheses regarding the probability distribution that governs the given vector of observations,  $\mathbf{X}$ . Under hypothesis  $\mathcal{H}_0$ ,  $\mathbf{X}$  is governed by  $p(\cdot|\theta_0)$ , and under hypothesis  $\mathcal{H}_1$ ,  $\mathbf{X}$  is governed by  $p(\cdot|\theta_1)$ . Suppose also that the a priori probability of  $\mathcal{H}_0$  to be the actual underlying hypothesis is  $\Pr\{\mathcal{H}_0\} = q$ , where  $0 \leq q \leq 1$  is given, and so the a priori probability of  $\mathcal{H}_1$  is the complementary probability,  $\Pr\{\mathcal{H}_1\} = 1 - q$ . When  $q = \frac{1}{2}$ , we say that the priors are equal; otherwise, the priors are unequal. A decision rule  $\Omega = (\Omega_0, \Omega_1)$  is a partition of the observation space,  $\mathcal{X}^n$ , into two disjoint decision regions,  $\Omega_0 \subseteq \mathcal{X}^n$  and its complementary region,  $\Omega_1 = \Omega_0^c$ : Given that  $\mathbf{X} = x \in \Omega_0$ , we decide in favor of  $\mathcal{H}_0$ ; otherwise, we decide in favor of  $\mathcal{H}_1$ . The probability of error, associated with the hypotheses  $\mathcal{H}_0$  and  $\mathcal{H}_1$ , referring to test points  $\theta_0$  and  $\theta_1$ , respectively, with priors  $q$  and  $1 - q$ , respectively, when using the decision rule  $\Omega$ , is defined as

$$P_e(q, \theta_0, \theta_1, \Omega) = q \cdot \int_{\Omega_1} p(x|\theta_0)dx + (1 - q) \cdot \int_{\Omega_0} p(x|\theta_1)dx. \tag{5}$$

A well-known elementary result in decision theory asserts that the optimal decision rule,  $\Omega^* = (\Omega_0^*, \Omega_1^*)$ , in the sense of minimizing  $P_e(q, \theta_0, \theta_1, \Omega)$ , is given by

$$\Omega_0^* = \{x \in \mathcal{X}^n : q \cdot p(x|\theta_0) \geq (1 - q) \cdot p(x|\theta_1)\} \tag{6}$$

$$\Omega_1^* = \{x \in \mathcal{X}^n : q \cdot p(x|\theta_0) < (1 - q) \cdot p(x|\theta_1)\}, \tag{7}$$

where it should be pointed out that attributing the case of a tie,  $q \cdot p(x|\theta_0) = (1 - q) \cdot p(x|\theta_1)$ , to  $\Omega_0^*$  is completely arbitrary, and could have been attributed alternatively to  $\Omega_1^*$  without affecting the probability of error. Here and in the sequel, the minimum probability of error, associated with  $\Omega^*$ , namely,  $P_e(q, \theta_0, \theta_1, \Omega^*)$ , will be denoted more simply by  $P_e(q, \theta_0, \theta_1)$ . On substituting  $\Omega^*$  into the above definition of the probability of error, we obtain

$$\begin{aligned} P_e(q, \theta_0, \theta_1) &= q \cdot \int_{\{x: q \cdot p(x|\theta_0) < (1-q) \cdot p(x|\theta_1)\}} p(x|\theta_0) dx + \\ &\quad (1 - q) \cdot \int_{\{x: q \cdot p(x|\theta_0) \geq (1-q) \cdot p(x|\theta_1)\}} p(x|\theta_1) dx \\ &= \int_{\{x: q \cdot p(x|\theta_0) < (1-q) \cdot p(x|\theta_1)\}} q \cdot p(x|\theta_0) dx + \\ &\quad \int_{\{x: q \cdot p(x|\theta_0) \geq (1-q) \cdot p(x|\theta_1)\}} (1 - q) \cdot p(x|\theta_1) dx \tag{8} \\ &= \int_{\{x: q \cdot p(x|\theta_0) < (1-q) \cdot p(x|\theta_1)\}} \min\{q \cdot p(x|\theta_0), (1 - q) p(x|\theta_1)\} dx + \\ &\quad \int_{\{x: q \cdot p(x|\theta_0) \geq (1-q) \cdot p(x|\theta_1)\}} \min\{q \cdot p(x|\theta_0), (1 - q) \cdot p(x|\theta_1)\} dx \\ &= \int_{\mathcal{X}^n} \min\{q \cdot p(x|\theta_0), (1 - q) \cdot p(x|\theta_1)\} dx \\ &= 1 - \int_{\mathcal{X}^n} \max\{q \cdot p(x|\theta_0), (1 - q) \cdot p(x|\theta_1)\} dx. \end{aligned}$$

The expression of the second to the last line will appear in some of our lower bounds in the sequel and will be recognized and interpreted as  $P_e(q, \theta_0, \theta_1)$ . The expression of the last line is the one that extends to multiple hypothesis testing, as will be detailed below.

The auxiliary problem of binary hypothesis testing extends from two hypotheses to a general number,  $m$ , of hypotheses, associated with  $m$  test points,  $\theta_0, \theta_1, \dots, \theta_{m-1}$ , in the following manner. Under hypothesis  $\mathcal{H}_i$ , the observation vector  $\mathbf{X}$  is governed by  $p(\cdot|\theta_i)$  and the a priori probability of  $\mathcal{H}_i$  being the underlying true hypothesis,  $\Pr\{\mathcal{H}_i\}$ , is denoted  $q_i$ ,  $i = 0, 1, \dots, m - 1$ , where  $q_0, q_1, \dots, q_{m-1}$  are given non-negative numbers summing to unity. Here, a decision rule  $\Omega = (\Omega_0, \Omega_1, \dots, \Omega_{m-1})$  is a partition of  $\mathcal{X}^n$  into  $m$  disjoint regions such that if  $x \in \Omega_i$ , we decide in favor of  $\mathcal{H}_i$ ,  $i = 0, 1, \dots, m - 1$ . The probability of error associated with  $\Omega, \theta_0, \theta_1, \dots, \theta_{m-1}$  and  $\mathbf{q} = (q_0, q_1, \dots, q_{m-1})$  is defined as

$$P_e(\mathbf{q}, \theta_0, \theta_1, \dots, \theta_{m-1}, \Omega) = \sum_{i=0}^{m-1} q_i \int_{\Omega_i^c} p(x|\theta_i) dx, \tag{9}$$

where  $\Omega_i^c$  is complementary to  $\Omega_i$ . The optimal MAP decision rule  $\Omega^*$  selects the hypothesis  $\mathcal{H}_i$  whose index  $i$  maximizes the product  $q_i \cdot p(x|\theta_i)$  among all  $i \in \{0, 1, \dots, m - 1\}$ , for the given  $x$ , where ties are broken arbitrarily. The probability of error, associated with  $\Omega^*$ , denoted  $P_e(\mathbf{q}, \theta_0, \theta_1, \dots, \theta_{m-1})$ , is well known to be given by

$$P_e(\mathbf{q}, \theta_0, \theta_1, \dots, \theta_{m-1}) = 1 - \int_{\mathcal{X}^n} \max\{q_0 \cdot p(x|\theta_0), q_1 \cdot p(x|\theta_1), \dots, q_{m-1} \cdot p(x|\theta_{m-1})\} dx. \tag{10}$$

Note that for  $m > 2$ , this is different from the expression

$$\int_{\mathcal{X}^n} \min\{q_0 \cdot p(x|\theta_0), q_1 \cdot p(x|\theta_1), \dots, q_{m-1} \cdot p(x|\theta_{m-1})\} dx,$$

as the latter can be interpreted as the probability that the index  $i$  of the true hypothesis  $\mathcal{H}_i$  minimizes (rather than maximizes) the product  $q_i p(\mathbf{x}|\theta_i)$  over  $i \in \{0, 1, \dots, m - 1\}$  for the given  $\mathbf{x}$ . Imagine an observer that, upon observing a realization  $\mathbf{x}$  of the random vector  $\mathbf{X}$ , creates a list of indices  $\{i\}$  with the  $k$  largest values of  $q_i p(\mathbf{x}|\theta_i)$  for some  $k < m$ , and an error is defined as the event where the correct  $i$  is not in that list. This is referred to as a list error, which is a term borrowed from the fields of coded communication and information theory. The last expression is the probability of list error for  $k = m - 1$ . We will encounter this expression later in certain versions of our lower bound. This completes the background needed about hypothesis testing.

As described in Section 1, our objective in this work is to derive relatively simple and easily computable lower bounds to  $r(\theta)$ , which are as tight as possible. While many existing lower bounds in the literature are satisfactory in terms of yielding the correct rate of convergence,  $1/\zeta_n^*$ , here we wish to improve the bound on the constant factor,  $r(\theta)$ . Many of our examples involve numerical calculations which include optimization over auxiliary parameters and occasionally also numerical integrations. All these calculations were carried out using MATLAB R2019a (9.6.0.1072779) 64-bit (win64).

### 3. Lower Bounds for Convex Symmetric Loss Functions

As explained earlier, here and in Section 4, we consider a scalar parameter, namely,  $d = 1$ .

**Theorem 1.** *Let the assumptions of Section 2 be satisfied for  $d = 1$  and let  $\rho(\cdot)$  be a symmetric convex loss function. Then,*

$$R_n(\Theta) \geq \sup_{\theta_0, \theta_1 \in \Theta} \left\{ 2 \cdot \rho\left(\frac{\theta_1 - \theta_0}{2}\right) \cdot \sup_{0 \leq q \leq 1} P_e(q, \theta_0, \theta_1) \right\}, \tag{11}$$

where  $P_e(q, \theta_0, \theta_1)$  is defined as in Equation (8).

**Proof of Theorem 1.** For every  $\theta_0, \theta_1 \in \Theta$  and  $q \in [0, 1]$ ,

$$\begin{aligned} R_n(\Theta) &\geq q \mathbb{E}_{\theta_0} \{ \rho(g_n[\mathbf{X}] - \theta_0) \} + (1 - q) \mathbb{E}_{\theta_1} \{ \rho(g_n[\mathbf{X}] - \theta_1) \} \\ &\stackrel{(a)}{=} \int_{\mathbb{R}^n} \left[ q \cdot p(\mathbf{x}|\theta_0) \rho(g_n[\mathbf{x}] - \theta_0) + \right. \\ &\quad \left. (1 - q) \cdot p(\mathbf{x}|\theta_1) \rho(\theta_1 - g_n[\mathbf{x}]) \right] d\mathbf{x} \\ &\geq 2 \cdot \int_{\mathbb{R}^n} \min\{q \cdot p(\mathbf{x}|\theta_0), (1 - q) \cdot p(\mathbf{x}|\theta_1)\} \times \\ &\quad \left[ \frac{1}{2} \rho(g_n[\mathbf{x}] - \theta_0) + \frac{1}{2} \rho(\theta_1 - g_n[\mathbf{x}]) \right] d\mathbf{x} \tag{12} \\ &\stackrel{(b)}{\geq} 2 \cdot \int_{\mathbb{R}^n} \min\{q \cdot p(\mathbf{x}|\theta_0), (1 - q) \cdot p(\mathbf{x}|\theta_1)\} \times \\ &\quad \rho\left(\frac{g_n[\mathbf{x}] - \theta_0}{2} + \frac{\theta_1 - g_n[\mathbf{x}]}{2}\right) d\mathbf{x} \\ &= 2 \cdot \rho\left(\frac{\theta_1 - \theta_0}{2}\right) \cdot \int_{\mathbb{R}^n} \min\{q \cdot p(\mathbf{x}|\theta_0), (1 - q) \cdot p(\mathbf{x}|\theta_1)\} d\mathbf{x} \\ &= 2 \cdot \rho\left(\frac{\theta_1 - \theta_0}{2}\right) \cdot P_e(q, \theta_0, \theta_1), \end{aligned}$$

where (a) is due to the assumed symmetry of  $\rho(\cdot)$  and (b) is by its assumed convexity. Since the inequality,

$$R_n(\Theta) \geq 2 \cdot \rho\left(\frac{\theta_1 - \theta_0}{2}\right) \cdot P_e(q, \theta_0, \theta_1) \tag{13}$$

applies to every  $\theta_0, \theta_1 \in \Theta$  and  $q \in [0, 1]$ , it applies, in particular, also to the supremum over these auxiliary parameters. This completes the proof of Theorem 1.  $\square$

Before we proceed, two comments are in order:

1. Note that  $P_e(q, \theta_0, \theta_1)$  is a concave function of  $q$  for fixed  $(\theta_0, \theta_1)$ , as it can be presented as the minimum among a family of affine functions of  $q$ , given by  $\min_{\Omega} [q \cdot P(\Omega|\theta_0) + (1 - q) \cdot P(\Omega^c|\theta_1)]$ , where  $\Omega$  runs over all possible subsets of the observation space,  $\mathcal{X}^n$ . Another way to see why this is true is by observing that  $P_e(q, \theta_0, \theta_1)$  is given in the second to the last line of (8) by an integral, whose integrand,  $\min\{q \cdot p(x|\theta_0), (1 - q) \cdot p(x|\theta_1)\}$ , is concave in  $q$ . Clearly,  $P_e(q, \theta_0, \theta_1) = 0$  whenever  $q = 0$  or  $q = 1$ . Thus,  $P_e(q, \theta_0, \theta_1)$  is maximized by some  $q$  between 0 and 1. If  $P_e(q, \theta_0, \theta_1)$  is strictly concave in  $q$ , then the maximizing  $q$  is unique.
2. Note that the lower bound (11) is tighter than the lower bound of  $\rho(\frac{\theta_1 - \theta_0}{2})P_e(\frac{1}{2}, \theta_0, \theta_1)$ , which was obtained in Equations (6)–(9a) in [18], both because of the factor of 2 and because of the freedom to optimize  $q$  rather than setting  $q = 1/2$ . In a further development of [18] the factor of 2 was accomplished too, but at the price of assuming that the density of the estimation error is symmetric about the origin (see discussion after (10) therein), which limits the class of estimators to which the bound applies. The factor of 2 and the degree of freedom  $q$  are also the two ingredients that make the difference between (11) and the lower bound due to Le Cam [20] (see also [26,28]). In Chapter 2 in [26] Guntuboyina reviews standard bounding techniques, including those of Le Cam, Assouad, and Fano. In particular, in Example 2.3.2 therein, Guntuboyina presents a lower bound in terms of the error probability associated with general priors, given by  $\frac{\eta}{2} \cdot P_e(q, \theta_0, \theta_1)$ , where in the case of two hypotheses,  $\eta = \min_{\vartheta} \{\rho(\theta_0 - \vartheta) + \rho(\theta_1 - \vartheta)\}$ , in our notation. Now, if  $\rho$  is symmetric and monotonically non-decreasing in the absolute error, then the minimizing  $\vartheta$  is given by  $\frac{\theta_0 + \theta_1}{2}$ , which yields  $\frac{\eta}{2} = \rho(\frac{\theta_1 - \theta_0}{2})$  and so, again, the resulting bound is of the same form as (11) except that it lacks the prefactor of 2.

Our first example demonstrates Theorem 1 on a somewhat technical but simple model, with an emphasis on the point that the optimal  $q$  may differ from 1/2 and that it is therefore useful to maximize w.r.t.  $q$  in order to improve the bound relative to the choice  $q = 1/2$ .

**Example 1.** Let  $X$  be a random variable distributed exponentially according to

$$p(x|\theta) = \theta e^{-\theta x}, \quad x \geq 0, \tag{14}$$

and  $\Theta = \{1, 2\}$ , so that the only possibility to select two different values of  $\theta$  in the lower bound are  $\theta_0 = 1$  and  $\theta_1 = 2$ . In terms of the hypothesis testing problem pertaining to the lower bound, the likelihood ratio test (LRT) is by comparison of  $q e^{-x}$  to  $(1 - q) \cdot 2e^{-2x}$ . Now, if  $2(1 - q) \leq q$ , or equivalently,  $1 \geq q \geq \frac{2}{3}$ , the decision is always in favor of  $\mathcal{H}_0$ , and then  $P_e(q, 1, 2) = 1 - q$ . For  $0 \leq q < \frac{2}{3}$ , the optimal LRT compares  $X$  to  $x_0(q) = \ln \frac{2(1-q)}{q}$ . If  $X > x_0(q)$ , one decides in favor of  $\mathcal{H}_0$ ; otherwise, one decides in favor of  $\mathcal{H}_1$ . Thus,

$$\begin{aligned} P_e(q, 1, 2) &= q \int_0^{x_0(q)} e^{-x} dx + (1 - q) \int_{x_0(q)}^{\infty} 2e^{-2x} dx \\ &= q[1 - e^{-x_0(q)}] + (1 - q)e^{-2x_0(q)} \\ &= q \left[ 1 - \frac{q}{2(1 - q)} \right] + (1 - q) \left[ \frac{q}{2(1 - q)} \right]^2 \\ &= \frac{q(4 - 5q)}{4(1 - q)}. \end{aligned} \tag{15}$$

In summary,

$$P_e(q, 1, 2) = \begin{cases} \frac{q(4-5q)}{4(1-q)} & 0 \leq q < \frac{2}{3} \\ 1 - q & \frac{2}{3} \leq q \leq 1 \end{cases} \tag{16}$$

It turns out that for  $q = \frac{1}{2}$ ,  $P_e(\frac{1}{2}, 1, 2) = \frac{3}{8} = 0.375$ , whereas the maximum is 0.382, attained at  $q = 1 - \frac{\sqrt{5}}{5} = 0.5528$ . Thus,

$$R_n(\Theta) \geq 2 \cdot 0.382 \cdot \rho\left(\frac{2-1}{2}\right) = 0.764 \cdot \rho\left(\frac{1}{2}\right). \tag{17}$$

This concludes Example 1.

In the above example, we considered just one observation,  $n = 1$ . From now on, we will refer to the case where  $n \gg 1$ . In particular, the following simple corollary to Theorem 1 yields a local asymptotic minimax lower bound.

**Corollary 1.** For a given  $\theta \in \Theta$  and a constant  $s$ , let  $\{\zeta_n\}_{n \geq 1}$  denote a positive sequence tending to zero with the property that

$$\lim_{n \rightarrow \infty} \max_q P_e(q, \theta, \theta + 2s\zeta_n) \tag{18}$$

exists and is given by a strictly positive constant, which will be denoted by  $P_e^\infty(\theta, s)$ . Also, let

$$\omega(s) \triangleq \lim_{u \rightarrow 0} \frac{\rho(s \cdot u)}{\rho(u)}. \tag{19}$$

Then, the local asymptotic minimax performance w.r.t.  $\zeta_n = 1/\rho(\zeta_n)$  is lower bounded by

$$r(\theta) \geq \sup_{s \in \mathbb{R}} \{2\omega(s) \cdot P_e^\infty(\theta, s)\}. \tag{20}$$

Corollary 1 is readily obtained from Theorem 1 by substituting  $\theta_0 = \theta$  and  $\theta_1 = \theta + 2s\zeta_n$  in Equation (11), then multiplying both sides of the inequality by  $\zeta_n = 1/\rho(\zeta_n)$ , and finally, taking the limit inferior of both sides.

Next, we study a few examples of the use of Corollary 1. As in Example 1, we emphasize again in Example 2 below the importance of having the degree of freedom to maximize over the prior  $q$  rather than to fix  $q = \frac{1}{2}$ . Also, in all the examples that were examined, the rate of convergence,  $1/\zeta_n = \rho(\zeta_n)$ , is the same as the optimal rate of convergence,  $1/\zeta_n^*$ . In other words, it is tight in the sense that there exists an estimator (for example, the maximum likelihood estimator) for which  $R_n(\theta, g_n)$  tends to zero at the same rate. In some of these examples, we compare our lower bound to  $r(\theta)$  to those of earlier reported results on the same models.

**Example 2.** Let  $X_1, \dots, X_n$  be independently, identically distributed (i.i.d.) random variables, uniformly distributed in the range  $[0, \theta]$ . In the corresponding hypothesis testing problem of Theorem 1, the hypotheses are  $\theta = \theta_0$  and  $\theta = \theta_1 > \theta_0$  with priors,  $q$  and  $1 - q$ . There are two cases: If  $q/\theta_0^n < (1 - q)/\theta_1^n$ , or equivalently,  $q < \theta_0^n / (\theta_0^n + \theta_1^n)$ , one decides always in favor of  $\mathcal{H}_1$  and so, the probability of error is  $q$ . If, on the other hand,  $q/\theta_0^n > (1 - q)/\theta_1^n$ , namely,  $q > \theta_0^n / (\theta_0^n + \theta_1^n)$ , we decide in favor of  $\mathcal{H}_1$  whenever  $\max_i X_i > \theta_0$  and then an error occurs only if  $\mathcal{H}_1$  is true, yet  $\max_i X_i < \theta_0$ , which happens with probability  $(1 - q) \left(\frac{\theta_0}{\theta_1}\right)^n$ . Thus,

$$P_e(q, \theta_0, \theta_1) = \begin{cases} q & q < \frac{\theta_0^n}{\theta_0^n + \theta_1^n} \\ (1 - q) \left(\frac{\theta_0}{\theta_1}\right)^n & q \geq \frac{\theta_0^n}{\theta_0^n + \theta_1^n} \end{cases} = \min \left\{ q, (1 - q) \left(\frac{\theta_0}{\theta_1}\right)^n \right\}, \tag{21}$$



which is readily seen to be maximized by  $q = \frac{\theta_0^n}{\theta_0^n + \theta_1^n}$  and then

$$\max_q P_c(q, \theta_0, \theta_1) = \frac{\theta_0^n}{\theta_0^n + \theta_1^n}. \tag{22}$$

Now, to apply Corollary 1, we let  $\theta_0 = \theta$  and  $\theta_1 = \theta_0(1 + 2\sigma/n)$ , which amounts to  $s = \theta_0\sigma = \theta\sigma$  and  $\xi_n = 1/n$ . Then,

$$P_c^\infty(\theta, s) = \lim_{n \rightarrow \infty} \frac{1}{1 + [1 + 2s/(\theta n)]^n} = \frac{1}{1 + e^{2s/\theta}}. \tag{23}$$

In the case of the MSE criterion,  $\rho(\epsilon) = \epsilon^2$ , we have  $\omega(s) = s^2$ , and so,

$$r(\theta) \geq \sup_{s \geq 0} \frac{2s^2}{1 + e^{2s/\theta}} = \theta^2 \cdot \sup_{u \geq 0} \frac{u^2}{2(1 + e^u)} \approx 0.2414\theta^2 \tag{24}$$

w.r.t.  $\zeta_n = 1/\rho(\xi_n) = 1/(1/n)^2 = n^2$ . This bound will be further improved upon in Section 4.

If instead of maximizing w.r.t.  $q$ , we select  $q = 1/2$ , then

$$P_c\left(\frac{1}{2}, \theta, \theta\left(1 + \frac{2\sigma}{n}\right)\right) = \frac{1}{2} \cdot \left[\frac{\theta}{\theta(1 + 2\sigma/n)}\right]^n \rightarrow \frac{1}{2} \cdot e^{-2\sigma} = \frac{1}{2} \cdot e^{-2s/\theta}, \tag{25}$$

and then the resulting bound would become

$$r(\theta) \geq \sup_{s \geq 0} s^2 e^{-2s/\theta} = \theta^2 \cdot \sup_{u \geq 0} \frac{u^2 e^{-u}}{4} \approx 0.1353\theta^2 \tag{26}$$

w.r.t.  $\zeta_n = n^2$ . Therefore, the maximization over  $q$  plays an important role here in terms of tightening the lower bound to  $r(\theta)$ .

More generally, for  $\rho(\epsilon) = |\epsilon|^t$  ( $t \geq 1$ ),  $\omega(s) = |s|^t$ , and we obtain

$$r(\theta) \geq \theta^t \cdot \sup_{u \geq 0} \frac{2u^t}{1 + e^{2u}}, \tag{27}$$

w.r.t.  $\zeta_n = n^t$ , where the supremum, which is in fact a maximum, can always be calculated numerically for every given  $t$ . For large  $t$ , the maximizing  $u$  is approximately  $t/2$ , which yields

$$r(\theta) \geq \frac{(t\theta)^t}{2^{t-1}(1 + e^t)}. \tag{28}$$

On the other hand, for  $q = 1/2$ , we end up with

$$r(\theta) \geq \sup_{s \geq 0} s^t e^{-2s/\theta} = \left(\frac{t\theta}{2e}\right)^t. \tag{29}$$

For large  $t$ , the bound of  $q = 1/2$  is inferior to the bound with the optimal  $q$ , by a factor of about  $1/2$ . This concludes Example 2.

**Example 3.** Let  $X_1, X_2, \dots, X_n$  be i.i.d. random variables, uniformly distributed in the interval  $[\theta, \theta + 1]$ . For the hypothesis testing problem, let  $\theta_1$  be chosen between  $\theta_0$  and  $\theta_0 + 1$ . Clearly, if  $\min_i X_i < \theta_1$ , the underlying hypothesis is certainly  $\mathcal{H}_1$ . Likewise, if  $\max_i X_i > \theta_0 + 1$ , the decision is in favor of  $\mathcal{H}_0$  with certainty. Thus, an error can occur only if all  $\{X_i\}$  fall in the interval  $[\theta_1, \theta_0 + 1]$ , an event that occurs with probability  $(\theta_0 + 1 - \theta_1)^n$ . In this event, the best to do is to select the hypothesis with the larger prior with a probability of error given by  $\min\{q, 1 - q\}$ . Thus,

$$P_c(q, \theta_0, \theta_1) = (\theta_0 + 1 - \theta_1)^n \cdot \min\{q, 1 - q\}, \tag{30}$$

and so,

$$\max_q P_c(q, \theta_0, \theta_1) = \frac{1}{2} [1 - (\theta_1 - \theta_0)]^n, \tag{31}$$

achieved by  $q = 1/2$ . Now, let us select  $\zeta_n = 1/n$ , which yields

$$\lim_{n \rightarrow \infty} \max_q P_c(q, \theta_0, \theta_1) = \lim_{n \rightarrow \infty} \frac{1}{2} \cdot \left(1 - \frac{2s}{n}\right)^n = \frac{1}{2} \cdot e^{-2s}. \tag{32}$$

For  $\rho(\varepsilon) = |\varepsilon|^t, (t \geq 1)$ , we have

$$r(\theta) \geq \sup_{s \geq 0} 2s^t \cdot \frac{1}{2} e^{-2s} = \sup_{s \geq 0} s^t e^{-2s} = \left(\frac{t}{2e}\right)^t \tag{33}$$

w.r.t.  $\zeta_n = 1/|1/n|^t = n^t$ . For the case of MSE,  $t = 2, r(\theta) \geq e^{-2} \approx 0.1353$ . The constant 0.1353 should be compared with  $\frac{1-1/\sqrt{2}}{128} \approx 0.0023$  (see Example 4.9 in [35]), which is two orders of magnitude smaller. This concludes Example 3.

**Example 4.** Let  $X_i = \theta + Z_i$ , where  $\{Z_i\}$  are i.i.d., Gaussian random variables with zero mean and variance  $\sigma^2$ . Here, for the corresponding binary hypothesis testing problem, the optimal value of  $q$  is always  $q^* = \frac{1}{2}$ . This can be readily seen from the concavity of  $P_c(q, \theta_0, \theta_1)$  in  $q$  and its symmetry around  $q = 1/2$ , as  $P_c(q, \theta_0, \theta_1) = P_c(1 - q, \theta_0, \theta_1)$ . Since

$$\begin{aligned} P_c\left(\frac{1}{2}, \theta_0, \theta_1\right) &= \Pr\left\{\sum_{i=1}^n Z_i \geq \frac{n(\theta_1 - \theta_0)}{2}\right\} \\ &= \Pr\left\{\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^n Z_i \geq \frac{\sqrt{n}(\theta_1 - \theta_0)}{2\sigma}\right\} \\ &= Q\left(\frac{\sqrt{n}(\theta_1 - \theta_0)}{2\sigma}\right), \end{aligned} \tag{34}$$

where

$$Q(t) \triangleq \int_t^\infty \frac{e^{-u^2/2} du}{\sqrt{2\pi}}, \tag{35}$$

we select  $\zeta_n = \frac{1}{\sqrt{n}}$ , which yields  $\theta_1 - \theta_0 = \frac{2s}{\sqrt{n}}$

$$P_c^\infty(\theta, s) = Q\left(\frac{s}{\sigma}\right), \tag{36}$$

and then for the MSE case,  $\omega(s) = s^2$ ,

$$r(\theta) \geq \sup_{s \geq 0} \left\{2s^2 Q\left(\frac{s}{\sigma}\right)\right\} = \sigma^2 \cdot \sup_{u \geq 0} \{2u^2 Q(u)\} \approx 0.3314\sigma^2 \tag{37}$$

w.r.t.  $\zeta_n = 1/(1/\sqrt{n})^2 = n$ , and so, the asymptotic lower bound to  $R_n(\theta, g_n)$  is  $0.3314\sigma^2/n$ .

We now compare this bound (which will be further improved in Section 4) with a few earlier reported results. In one of the versions of Le Cam's bound (see Example 4.7 in [35]) for the same model, the lower bound to  $r(\theta)$  turns out to be  $\frac{\sigma^2}{24} = 0.0417\sigma^2$ , namely, an order of magnitude smaller. Also, in Example 3.1 in [28], another version of Le Cam's method yields  $r(\theta) \geq (1 - \sqrt{1/2})\sigma^2/8 = 0.0366\sigma^2$ . According to Corollary 4.3 in [36],  $r(\theta) \geq \frac{\sigma^2}{8e} = 0.046\sigma^2$ . Yet another comparison is with Theorem 5.9 in [37], where we find an inequality, which in our notation reads as follows:

$$\sup_{\theta \in \Theta} P_\theta \left\{ (g_n[\mathbf{X}] - \theta)^2 \geq \frac{2\alpha\sigma^2}{n} \right\} \geq \frac{1}{2} - \alpha \quad \alpha \in \left(0, \frac{1}{2}\right). \tag{38}$$

Combining it with Chebychev’s inequality yields

$$\sup_{\theta \in \Theta} \mathbb{E}_{\theta}(g_n[\mathbf{X}] - \theta)^2 \geq \frac{\sigma^2}{n} \cdot \max_{0 \leq \alpha \leq 1/2} \alpha(1 - 2\alpha) = \frac{0.125\sigma^2}{n}. \tag{39}$$

In [23] (p. 257), it is shown that when  $\Theta = \mathbb{R}$ , the minimax estimator for this model is the sample mean, and so, in this case, the correct constant in front of  $\sigma^2$  is actually 1. This concludes Example 4.

The next example is related to Example 4, as it is based on the use of the central limit theorem (CLT), which means that the Gaussian tail distribution is used here too.

**Example 5.** Consider an exponential family,

$$p(\mathbf{x}|\theta) = \prod_{i=1}^n p(x_i|\theta) = \prod_{i=1}^n \frac{e^{\theta T(x_i)}}{Z(\theta)} = \frac{\exp\{\theta \sum_{i=1}^n T(x_i)\}}{Z^n(\theta)}. \tag{40}$$

where  $T(\cdot)$  is a given function and  $Z(\theta)$  is a normalization function given by

$$Z(\theta) = \int_{\mathbb{R}} e^{\theta T(x)} dx, \tag{41}$$

assuming that the integral converges. In the auxiliary binary hypothesis problem, the test statistic is  $\sum_{i=1}^n T(X_i)$ . If  $q = 1/2$ ,  $\xi_n = 1/\sqrt{n}$  and  $\theta_1 - \theta_0 = 2s/\sqrt{n}$ , the LRT amounts to examining whether  $\sum_{i=1}^n [T(X_i) - \mathbb{E}_{\theta}\{T(X_i)\}]$  is larger than

$$\frac{s}{\sqrt{n}} \cdot n \frac{d^2 \ln Z(\theta)}{d\theta^2} = s\sqrt{n} \cdot \frac{d^2 \ln Z(\theta)}{d\theta^2}.$$

In this case, the probability of error can be asymptotically assessed using the CLT, which after a simple algebraic manipulation, becomes:

$$\begin{aligned} P_e^{\infty}(\theta, s) &= Q\left(\frac{sd^2 \ln Z(\theta)/d\theta^2}{\sqrt{d^2 \ln Z(\theta)/d\theta^2}}\right) \\ &= Q\left(s\sqrt{\frac{d^2 \ln Z(\theta)}{d\theta^2}}\right) \\ &= Q\left(s\sqrt{I(\theta)}\right), \end{aligned} \tag{42}$$

where  $I(\theta)$  is the Fisher information. Thus, for the MSE,

$$r(\theta) \geq \sup_{s \geq 0} 2s^2 Q\left(s\sqrt{I(\theta)}\right) \approx \frac{0.3314}{I(\theta)} \tag{43}$$

w.r.t.  $\zeta_n = 1/(1/\sqrt{n})^2 = n$ . This concludes Example 5.

In several earlier examples, our bound was shown to outperform (sometimes significantly so) earlier reported bounds for the corresponding models. However, to be honest and fair, we should not ignore the fact that there are also situations where our bound may not be tighter than earlier bounds applied to the same model. Such a case is demonstrated in Example 6 below, where our result is compared to those of Ziv and Zakai [18] and Chazan, Zakai, and Ziv [4] in the context of estimating the delay of a known continuous-time signal corrupted by additive white Gaussian noise (for further developments and applications of the Ziv–Zakai and the Chazan–Zakai–Ziv bounds, see, e.g., [38–45] and references therein). Having said that, it should also be kept in mind that our emphasis in this work is on

bounds that are relatively simple and easy to calculate (at least numerically), whereas the Chazan–Zakai–Ziv bound, although very tight in many situations, is notoriously difficult to calculate. Indeed, in this setting, the explicit behavior of the resulting complicated bound of [4] is clearly transparent only at high values of the signal-to-noise ratio (SNR)—see Equations (11), (12), and (14) in [4].

**Example 6.** Let  $X(t) = s(t, \theta) + N(t)$ ,  $t \in [0, T]$ , where  $N(t)$  is additive white Gaussian noise (AWGN) with double-sided spectral density  $N_0$  and  $s(t, \theta)$  is a deterministic signal that depends on the unknown parameter,  $\theta$ . It is assumed that the signal energy,  $E = \int_0^T s^2(t, \theta) dt$ , does not depend on  $\theta$  (which is the case, for example, when  $\theta$  is a delay parameter of a pulse fully contained in the observation interval, or when  $\theta$  is the frequency or the phase of a sinusoidal waveform). We further assume that  $s(t, \theta)$  is at least twice differentiable w.r.t.  $\theta$ , and that the energies of the first two derivatives are also independent of  $\theta$ . Then, as shown in Appendix A, for small  $|\theta_1 - \theta_0|$ ,

$$\begin{aligned} \varrho(\theta_0, \theta_1) &\triangleq \frac{1}{E} \int_0^T s(t, \theta_0) s(t, \theta_1) dt \\ &= 1 - \frac{(\theta_1 - \theta_0)^2}{2E} \int_0^T \left[ \frac{\partial s(t, \theta)}{\partial \theta} \Big|_{\theta=\theta_0} \right]^2 dt + o(|\theta_1 - \theta_0|^2) \\ &= 1 - \frac{(\theta_1 - \theta_0)^2 \dot{E}}{2E} + o(|\theta_1 - \theta_0|^2), \end{aligned} \tag{44}$$

where  $\dot{E}$  is the energy  $\dot{s}(t, \theta) = \partial s(t, \theta) / \partial \theta$ .

The optimal LRT in deciding between the two hypotheses is based on the comparison between the correlations,  $\int_0^T X(t) s(t, \theta_0) dt$  and  $\int_0^T X(t) s(t, \theta_1) dt$ . Again, the optimal value of  $q$  is  $q^* = 1/2$ . Thus,

$$\begin{aligned} P_e\left(\frac{1}{2}, \theta_0, \theta_1\right) &= Q\left(\sqrt{\frac{E}{2N_0}} [1 - \varrho(\theta_0, \theta_1)]\right) \\ &\approx Q\left(\sqrt{\frac{E \dot{E} (\theta_1 - \theta_0)^2}{4EN_0}}\right) \\ &= Q\left(\sqrt{\frac{\dot{E}}{4N_0}} \cdot |\theta_1 - \theta_0|\right) \\ &= Q\left(\sqrt{\frac{\dot{P}}{4N_0}} \cdot \sqrt{T} |\theta_1 - \theta_0|\right), \end{aligned} \tag{45}$$

where  $\dot{P} = \dot{E} / T$  is the power of  $\dot{s}(t, \theta)$ . Since we are dealing here with continuous time, instead of a sequence  $\xi_n$ , we use a function,  $\xi(T)$ , of the observation time,  $T$ , which in this case would be  $\xi(T) = \frac{1}{\sqrt{T}}$ . Let  $\theta_0 = \theta$  and  $\theta_1 = \theta + \frac{2s}{\sqrt{T}}$ . Then,

$$P_e^\infty(\theta, s) = Q\left(\sqrt{\frac{\dot{P}}{N_0}} \cdot |s|\right), \tag{46}$$

which, for the MSE case, yields

$$\begin{aligned} r(\theta) &\geq \sup_{s \geq 0} \left\{ 2s^2 Q\left(\sqrt{\frac{\dot{P}}{N_0}} \cdot s\right) \right\} \\ &= \frac{N_0}{\dot{P}} \sup_{u \geq 0} \{ 2u^2 Q(u) \} \\ &\approx \frac{0.3314 N_0}{\dot{P}}. \end{aligned} \tag{47}$$

w.r.t.  $\zeta(T) = 1/(1/\sqrt{T})^2 = T$ , which means that the minimax loss is lower bounded by  $r(\theta)/T \geq 0.3314N_0/\dot{E}$ . This has the same form as the Cramér–Rao lower bound (CRLB), except that the multiplicative factor is 0.3314 rather than 1. In Equation (20) in [18], the bound is of the same form, but with a multiplicative constant of 0.16 for a high signal-to-noise ratio (SNR). However, in [4], the constant of proportionality was improved to 1 in the high SNR limit, just like in the Cramér–Rao lower bound for the same model. The constant 0.3314 will be improved later to 0.4549 (under Example 9 in the sequel), but it will still be below 1.

The case where  $s(t, \theta)$  is not everywhere differentiable w.r.t.  $\theta$  can be handled in a similar manner, but some caution should be exercised. For example, consider the model,

$$X(t) = s(t - \theta) + N(t), \tag{48}$$

where  $-\infty < t < \infty$ ,  $\theta \in [0, T]$ ,  $N(t)$  is AWGN as before, and  $s(\cdot)$  is a rectangular pulse with duration  $\Delta$  and amplitude  $\sqrt{E/\Delta}$ ,  $E$  being the signal energy. Here,  $q(\theta, \theta + \delta) = 1 - |\delta|/\Delta$ ; namely, it also includes a linear term in  $|\delta|$ , not just the quadratic one. This changes the asymptotic behavior of the resulting lower bound to  $r(\theta)$ , which turns out to be  $0.7544(N_0\Delta/P)^2$  w.r.t.  $T^2$  (namely, a minimax lower bound of  $0.7544(N_0\Delta/E)^2$ ). It is interesting to compare this bound to the Chapman–Robbins bound for the same model, which is a local bound of the same form but with a multiplicative constant of 0.1602 instead of 0.7544, and which is limited to unbiased estimators. The Chazan–Zakai–Ziv bound [4] for this case is difficult to calculate, but in the high SNR regime, it behaves like  $3(N_0\Delta/P)^2$  w.r.t.  $T^2$ .

It is conceivable that the Chazan–Zakai–Ziv bound for estimating the delay of non-differential signals in Gaussian white noise has been improved even further ever since its publication in 1975, but the point of this particular example remains: our bound may not always be the best available bound. Still, since our bound is easy to calculate, it is worth comparing it to other bounds for every given model. This concludes Example 6.

#### 4. Bounds Based on the Minimum Expected Loss over Some Test Points

In this section, we derive our second family of bounds for the scalar case,  $d = 1$ .

##### 4.1. Two Test Points

The following generic, yet conceptually very simple, lower bound assumes neither symmetry nor convexity of the loss function  $\rho(\cdot)$ . For a given  $(\mathbf{x}, q, \theta_0, \theta_1) \in \mathbb{R}^n \times [0, 1] \times \Theta^2$ , let us define

$$\psi(\mathbf{x}, q, \theta_0, \theta_1) = \min_{\vartheta} \{qp(\mathbf{x}|\theta_0)\rho(\vartheta - \theta_0) + (1 - q)p(\mathbf{x}|\theta_1)\rho(\vartheta - \theta_1)\}. \tag{49}$$

Then,

$$\begin{aligned} R_n(\Theta) &\geq \sup_{\theta_0, \theta_1, q} q\mathbb{E}_{\theta_0} \{\rho(g_n[\mathbf{X}] - \theta_0)\} + (1 - q)\mathbb{E}_{\theta_1} \{\rho(g_n[\mathbf{X}] - \theta_1)\} \\ &= \sup_{\theta_0, \theta_1, q} \int_{\mathbb{R}^n} [qp(\mathbf{x}|\theta_0)\rho(g_n[\mathbf{x}] - \theta_0) + (1 - q)p(\mathbf{x}|\theta_1)\rho(g_n[\mathbf{x}] - \theta_1)]d\mathbf{x} \tag{50} \\ &\geq \sup_{\theta_0, \theta_1, q} \int_{\mathbb{R}^n} \psi(\mathbf{x}, q, \theta_0, \theta_1)d\mathbf{x}. \end{aligned}$$

If we further assume the symmetry of  $\rho$ , then it is easy to see that the minimizer  $\vartheta^*$ , which achieves  $\psi(\mathbf{x}, q, \theta_0, \theta_1)$ , is always within the interval  $[\theta_0, \theta_1]$ . This is because the objective increases monotonically as we move away from the interval  $[\theta_0, \theta_1]$  in either direction. Of course, this simple idea can easily be extended to apply to the weighted sums of more than two points, in principle, but it would become more complicated—see the next subsection for three such points.

If  $\rho$  is concave, then the minimizing  $\vartheta$  is either  $\theta_0$  or  $\theta_1$ , depending on which of  $qp(x|\theta_0)$  and  $(1 - q)p(x|\theta_1)$  is smaller, and the bound becomes

$$R_n(\Theta) \geq \sup_{\theta_0, \theta_1, q} \rho(\theta_1 - \theta_0) P_\epsilon(q, \theta_0, \theta_1). \tag{51}$$

Minimality at the edge points may also happen for some loss functions that are not concave, like the loss function  $\rho(u) = 1\{|u| \geq \Delta\}$ .

The generic lower bound (50) is more general than our first bound in the sense that it does not require convexity or symmetry of  $\rho$ , but the downside is that the resulting expressions are harder to deal with directly, as will be seen shortly. For loss functions other than the MSE or general moments of the estimation error, it may not be a trivial task even to derive a closed-form expression of  $\psi(x, q, \theta_0, \theta_1)$  (i.e., to carry out the minimization associated with the definition of  $\psi$ ).

For the case of the MSE,  $\rho(\epsilon) = \epsilon^2$ , the calculation of  $\psi(x, q, \theta_0, \theta_1)$  is straightforward, and it readily yields

$$\psi(x, q, \theta_0, \theta_1) = (\theta_1 - \theta_0)^2 \cdot \frac{qp(x|\theta_0) \cdot (1 - q)p(x|\theta_1)}{qp(x|\theta_0) + (1 - q)p(x|\theta_1)}. \tag{52}$$

However, it may not be convenient to integrate this function of  $x$  due to the summation at the denominator. One way to alleviate this difficulty is to observe that

$$\begin{aligned} \psi(x, q, \theta_0, \theta_1) &\geq (\theta_1 - \theta_0)^2 \cdot \frac{qp(x|\theta_0) \cdot (1 - q)p(x|\theta_1)}{2 \max\{qp(x|\theta_0), (1 - q)p(x|\theta_1)\}} \\ &= \frac{1}{2} \cdot (\theta_1 - \theta_0)^2 \cdot \min\{qp(x|\theta_0), (1 - q)p(x|\theta_1)\}, \end{aligned} \tag{53}$$

which after integration yields again

$$R_n(\Theta) \geq \sup_{\theta_0, \theta_1, q} \left\{ \frac{1}{2} (\theta_1 - \theta_0)^2 \cdot P_\epsilon(q, \theta_0, \theta_1) \right\}, \tag{54}$$

exactly as in Theorem 1 in the special case of the MSE. This indicates that the bound (50) is at least as tight as the bound of Theorem 1 for the MSE.

It turns out, however, that we can do better than bounding the denominator,  $qp(x|\theta_0) + (1 - q)p(x|\theta_1)$ , by  $2 \cdot \max\{qp(x|\theta_0), (1 - q)p(x|\theta_1)\}$  for the purpose of obtaining a more convenient integrand. To this end, we invoke the following lemma, whose proof appears in Appendix B.

**Lemma 1.** *Let  $k$  be a positive integer and let  $a_1, \dots, a_k$  be positive reals. Then,*

$$\sum_{i=1}^k a_i = \inf_{(r_1, \dots, r_k) \in \mathcal{S}} \max \left\{ \frac{a_1}{r_1}, \dots, \frac{a_k}{r_k} \right\}, \tag{55}$$

where  $\mathcal{S}$  is the interior of the  $k$ -dimensional simplex, namely, the set of all vectors  $(r_1, \dots, r_k)$  with strictly positive components that sum to unity.

Applying Lemma 1 with the assignments  $k = 2$ ,  $a_1 = qp(x|\theta_0)$ , and  $a_2 = (1 - q)p(x|\theta_1)$ , we have

$$qp(x|\theta_0) + (1 - q)p(x|\theta_1) = \inf_{r \in (0,1)} \max \left\{ \frac{qp(x|\theta_0)}{r}, \frac{(1 - q)p(x|\theta_1)}{1 - r} \right\}. \tag{56}$$

Thus,

$$\psi(x, q, \theta_0, \theta_1) = \frac{qp(x|\theta_0) \cdot (1 - q)p(x|\theta_1)}{\inf_{r \in (0,1)} \max\{qp(x|\theta_0)/r, (1 - q)p(x|\theta_1)/(1 - r)\}}$$

$$\begin{aligned}
 &= \sup_{r \in (0,1)} \frac{qp(x|\theta_0) \cdot (1-q)p(x|\theta_1)}{\max\{qp(x|\theta_0)/r, (1-q)p(x|\theta_1)/(1-r)\}} \tag{57} \\
 &= \sup_{r \in (0,1)} \begin{cases} r(1-q)p(x|\theta_1) & r \leq r^* \\ (1-r)qp(x|\theta_0) & r \geq r^* \end{cases} \\
 &= \sup_{r \in (0,1)} \min\{r(1-q)p(x|\theta_1), (1-r)qp(x|\theta_0)\},
 \end{aligned}$$

where

$$r^* = \frac{qp(x|\theta_0)}{qp(x|\theta_0) + (1-q)p(x|\theta_1)}. \tag{58}$$

Thus, the bound becomes

$$\begin{aligned}
 R_n(\Theta) &\geq \sup_{\{(\theta_0, \theta_1, q) \in \Theta^2 \times (0,1)\}} (\theta_1 - \theta_0)^2 \times \\
 &\quad \int_{\mathbb{R}^n} \sup_{r \in (0,1)} \min\{r(1-q)p(x|\theta_1), (1-r)qp(x|\theta_0)\} dx \\
 &\geq \sup_{(\theta_0, \theta_1, q) \in \Theta^2 \times (0,1)} \sup_{r \in (0,1)} (\theta_1 - \theta_0)^2 \times \\
 &\quad \int_{\mathbb{R}^n} \min\{r(1-q)p(x|\theta_1), (1-r)qp(x|\theta_0)\} dx \\
 &= \sup_{(\theta_0, \theta_1, q, r) \in \Theta^2 \times (0,1)^2} (\theta_1 - \theta_0)^2 \cdot (q + r - 2qr) \times \\
 &\quad \int_{\mathbb{R}^n} \min\left\{ \frac{r(1-q)}{q + r - 2qr} \cdot p(x|\theta_1), \frac{(1-r)q}{q + r - 2qr} \cdot p(x|\theta_0) \right\} dx \\
 &= \sup_{(\theta_0, \theta_1, q, r) \in \Theta^2 \times (0,1)^2} (\theta_1 - \theta_0)^2 \cdot (q + r - 2qr) \cdot P_e\left(\frac{(1-r)q}{q + r - 2qr}, \theta_0, \theta_1\right).
 \end{aligned}$$

The bound of Theorem 1 for the MSE is obtained as a special case of  $r = 1/2$ . Therefore, after the optimization over the additional degree of freedom,  $r$ , the resulting bound cannot be worse than the MSE bound of Theorem 1. In fact, it may strictly improve as we will demonstrate shortly. The choice  $r = q$  gives a prior of 1/2 in the error probability factor, and then the maximum of the external factor,  $q + r - 2rq = 2q(1 - q)$ , is maximized by  $q = 1/2$ .

**Example 7.** To demonstrate the new bound for the MSE, let us revisit Example 2 and see how it improves the multiplicative constant. In that example,

$$P_e\left(\frac{(1-r)q}{q+r-2qr}, \theta_0, \theta_1\right) = \min\left\{ \frac{(1-r)q}{q+r-2qr}, \frac{r(1-q)}{q+r-2qr} \cdot \left(\frac{\theta_0}{\theta_1}\right)^n \right\}. \tag{59}$$

Let us denote  $\alpha = (\theta_0/\theta_1)^n$  and recall that  $\alpha \in (0, 1)$ , provided that we select  $\theta_1 > \theta_0$ . Then,

$$(q + r - 2qr) \cdot P_e\left(\frac{(1-r)q}{q+r-2qr}, \theta_0, \theta_1\right) = \min\{(1-r)q, r(1-q)\alpha\}. \tag{60}$$

The maximum w.r.t.  $q$  is attained when  $(1-r)q = r(1-q)\alpha$ , namely, for  $q = q^* \triangleq \alpha r / [1 - (1 - \alpha)r]$ , which yields

$$\max_q (q + r - 2qr) \cdot P_e\left(\frac{(1-r)q}{q+r-2qr}, \theta_0, \theta_1\right) = \frac{\alpha r(1-r)}{1 - (1 - \alpha)r}. \tag{61}$$

Let us denote  $\beta \triangleq 1 - \alpha \in (0, 1)$ . The maximum of  $r(1 - r)/(1 - \beta r)$  is attained for

$$r = r^* \triangleq \frac{1 - \sqrt{1 - \beta}}{\beta} = \frac{1 - \sqrt{\alpha}}{1 - \alpha} = \frac{1}{1 + \sqrt{\alpha}}, \tag{62}$$

which yields

$$\begin{aligned} \max_{q,r} (q + r - 2qr) \cdot P_e \left( \frac{(1 - r)q}{q + r - 2qr}, \theta_0, \theta_1 \right) &= \sup_{r \in (0,1)} \frac{\alpha r(1 - r)}{1 - (1 - \alpha)r} \\ &= \alpha \cdot \left( \frac{1 - \sqrt{\alpha}}{1 - \alpha} \right)^2 \\ &= \frac{\alpha}{(1 + \sqrt{\alpha})^2}. \end{aligned} \tag{63}$$

To obtain a local bound in the spirit of Corollary 1, take  $\theta_0 = \theta$ ,  $\theta_1 = \theta(1 + \frac{s}{n\theta})$ , which yields  $\alpha = e^{-s/\theta}$  in the limit of large  $n$ , and so,

$$r(\theta) \geq \sup_{s \geq 0} \frac{s^2 e^{-s/\theta}}{(1 + e^{-s/[2\theta]})^2} = \theta^2 \cdot \sup_{u \geq 0} u^2 \cdot \frac{e^{-u}}{(1 + e^{-u/2})^2} \approx 0.3102\theta^2, \tag{64}$$

w.r.t.  $\zeta_n = n^2$ , which improves on our earlier bound in Example 2,  $r(\theta) \geq 0.2414\theta^2$ . This concludes Example 7.

More generally, for general moments of the estimation error, a similar derivation yields the following:

**Theorem 2.** For  $\rho(\varepsilon) = |\varepsilon|^t, t \geq 1$  (not necessarily an integer),

$$R_n(\Theta) \geq \sup_{\theta_0, \theta_1, q, r} |\theta_1 - \theta_0|^t [(1 - r)^{t-1}q + r^{t-1}(1 - q)] \cdot P_e \left( \frac{(1 - r)^{t-1}q}{(1 - r)^{t-1}q + r^{t-1}(1 - q)}, \theta_0, \theta_1 \right). \tag{65}$$

Applying the local version of Theorem 2 to Example 7, we obtain

$$r(\theta) \geq \theta^t \cdot \sup_{s > 0} \frac{s^t e^{-s}}{(1 + e^{-s/t})^t} \tag{66}$$

w.r.t.  $\zeta_n = n^t$ . Changing the optimization variable from  $s$  to  $\sigma = s/t$ , we end up with

$$r(\theta) \geq (\theta t)^t \cdot \left[ \sup_{\sigma > 0} \frac{\sigma}{e^\sigma + 1} \right]^t \approx (0.2785t\theta)^t. \tag{67}$$

The factor of  $(0.2785t)^t$  should be compared with  $(1/2e)^t = 0.1839^t$  of Example 2, pertaining to the choice  $q = r = 1/2$ . The gap increases exponentially with  $t$ . For the maximum likelihood estimator pertaining to Example 2, which is  $g_n[\mathbf{X}] = \max_i X_i$ , it is easy to show that whenever  $t$  is an integer,

$$\mathbb{E}_\theta \{ |\hat{\theta} - \theta|^t \} = \frac{n!t!\theta^t}{(n + t)!} = \frac{t!\theta^t}{(n + 1)(n + 2) \cdots (n + t)}, \tag{68}$$

and so, the asymptotic gap is between  $t!$  and  $(0.2785t)^t$ . Considering the Stirling approximation, the ratio between the upper bound and the lower bound is about  $\sqrt{2\pi t} \cdot (1.3211)^t$ .

#### 4.2. Three Test Points

The idea behind the bounds of Section 4.1 can be conceptually extended to be based on more than two test points, but the resulting expressions become cumbersome very quickly as the number of test points grows. For three points, however, this is still manageable and



can provide improved bounds. Let us select the three test points to be  $\theta_0 - \Delta$ ,  $\theta_0$ , and  $\theta_0 + \Delta$  for some  $\theta_0$  and  $\Delta$ , and let us assign weights  $q$ ,  $r$ , and  $w = 1 - q - r$ . Consider the bound

$$\begin{aligned}
 R_n(\Theta) &\geq q\mathbb{E}_{\theta_0-\Delta}\{\rho(g_n[\mathbf{X}] - \theta_0 + \Delta)\} + r\mathbb{E}_{\theta_0}\{\rho(g_n[\mathbf{X}] - \theta_0)\} + \\
 &\quad w\mathbb{E}_{\theta_0+\Delta}\{\rho(g_n[\mathbf{X}] - \theta_0 - \Delta)\} \\
 &= \int_{\mathbb{R}^n} \left[ q \cdot p(\mathbf{x}|\theta_0 - \Delta)\rho(g_n[\mathbf{x}] - \theta_0 + \Delta) + r \cdot p(\mathbf{x}|\theta_0)\rho(g_n[\mathbf{x}] - \theta_0) + \right. \\
 &\quad \left. w \cdot p(\mathbf{x}|\theta_0 + \Delta)\rho(g_n[\mathbf{x}] - \theta_0 - \Delta) \right] d\mathbf{x} \tag{69} \\
 &\geq \int_{\mathbb{R}^n} \Psi(\mathbf{x}, \theta_0, \Delta, q, r) d\mathbf{x},
 \end{aligned}$$

where

$$\begin{aligned}
 \Psi(\mathbf{x}, \theta_0, \Delta, q, r) &= \min_{\vartheta} \left\{ q \cdot p(\mathbf{x}|\theta_0 - \Delta)\rho(\vartheta - \theta_0 + \Delta) + r \cdot p(\mathbf{x}|\theta_0)\rho(\vartheta - \theta_0) + \right. \\
 &\quad \left. w \cdot p(\mathbf{x}|\theta_0 + \Delta)\rho(\vartheta - \theta_0 - \Delta) \right\}. \tag{70}
 \end{aligned}$$

Considering the case of the MSE,

$$\begin{aligned}
 \Psi(\mathbf{x}, \theta_0, \Delta, q, r) &= \min_{\vartheta} \left\{ q \cdot p(\mathbf{x}|\theta_0 - \Delta)(\vartheta - \theta_0 + \Delta)^2 + r \cdot p(\mathbf{x}|\theta_0)(\vartheta - \theta_0)^2 + \right. \\
 &\quad \left. w \cdot p(\mathbf{x}|\theta_0 + \Delta)(\vartheta - \theta_0 - \Delta)^2 \right\} \tag{71}
 \end{aligned}$$

can be found in closed-form. Denoting temporarily  $a = qp(\mathbf{x}|\theta_0 - \Delta)$ ,  $b = rp(\mathbf{x}|\theta_0)$ , and  $c = wp(\mathbf{x}|\theta_0 + \Delta)$ ,  $\Psi(\mathbf{x}, \theta_0, \Delta, q, r)$  is attained by

$$\vartheta = \vartheta^* = \frac{a(\theta_0 - \Delta) + b\theta_0 + c(\theta_0 + \Delta)}{a + b + c} = \theta_0 + \frac{(c - a)\Delta}{a + b + c}. \tag{72}$$

On substituting  $\vartheta^*$  into the sum of squares, we end up with

$$\begin{aligned}
 \Psi(\mathbf{x}, \theta_0, \Delta, q, r) &= \frac{[a(b + 2c)^2 + b(c - a)^2 + c(2a + b)^2]\Delta^2}{(a + b + c)^2} \\
 &= \frac{(ab + bc + 4ac)\Delta^2}{a + b + c} \\
 &= \frac{[qrp(\mathbf{x}|\theta_0 - \Delta)p(\mathbf{x}|\theta_0) + rwp(\mathbf{x}|\theta_0)p(\mathbf{x}|\theta_0 + \Delta) + 4qwp(\mathbf{x}|\theta_0 - \Delta)p(\mathbf{x}|\theta_0 + \Delta)]\Delta^2}{qp(\mathbf{x}|\theta_0 - \Delta) + rp(\mathbf{x}|\theta_0) + wp(\mathbf{x}|\theta_0 + \Delta)}.
 \end{aligned}$$

The lower bound on the MSE is then the sum of three integrals

$$I_1 = qr\Delta^2 \cdot \int_{\mathbb{R}^n} \frac{p(\mathbf{x}|\theta_0 - \Delta)p(\mathbf{x}|\theta_0)d\mathbf{x}}{qp(\mathbf{x}|\theta_0 - \Delta) + rp(\mathbf{x}|\theta_0) + wp(\mathbf{x}|\theta_0 + \Delta)} \tag{73}$$

$$I_2 = rw\Delta^2 \cdot \int_{\mathbb{R}^n} \frac{p(\mathbf{x}|\theta_0)p(\mathbf{x}|\theta_0 + \Delta)d\mathbf{x}}{qp(\mathbf{x}|\theta_0 - \Delta) + rp(\mathbf{x}|\theta_0) + wp(\mathbf{x}|\theta_0 + \Delta)} \tag{74}$$

$$I_3 = 4qw\Delta^2 \cdot \int_{\mathbb{R}^n} \frac{p(\mathbf{x}|\theta_0 - \Delta)p(\mathbf{x}|\theta_0 + \Delta)d\mathbf{x}}{qp(\mathbf{x}|\theta_0 - \Delta) + rp(\mathbf{x}|\theta_0) + wp(\mathbf{x}|\theta_0 + \Delta)}. \tag{75}$$

**Example 8.** Revisiting again Example 2, for a given  $t > 0$ , let us denote  $\mathcal{C}(t) = [0, t]^n$ , and then  $p(\mathbf{x}|\theta_0 + i\Delta) = [\theta_0 + i\Delta]^{-n} \cdot \mathcal{I}\{\mathbf{x} \in \mathcal{C}(\theta_0 + i\Delta), i = -1, 0, 1\}$ . Then,

$$I_1 = qr\Delta^2 \cdot \int_{\mathcal{C}(\theta_0-\Delta)} \frac{(\theta_0 - \Delta)^{-n}\theta_0^{-n}d\mathbf{x}}{q(\theta_0 - \Delta)^{-n} + r\theta_0^{-n} + w(\theta_0 + \Delta)^{-n}}$$

$$\begin{aligned}
 &= qr\Delta^2 \cdot \frac{(\theta_0 - \Delta)^n(\theta_0 - \Delta)^{-n}\theta_0^{-n}}{q(\theta_0 - \Delta)^{-n} + r\theta_0^{-n} + w(\theta_0 + \Delta)^{-n}} \\
 &= \frac{qr\Delta^2\theta_0^{-n}}{q(\theta_0 - \Delta)^{-n} + r\theta_0^{-n} + w(\theta_0 + \Delta)^{-n}} \\
 &= \frac{qr\Delta^2}{q[\theta_0/(\theta_0 - \Delta)]^n + r + w[\theta_0/(\theta_0 + \Delta)]^n}.
 \end{aligned} \tag{76}$$

For  $\Delta = \theta_0 s/n$ , we then have, for large  $n$ :

$$I_1 \sim \frac{\theta_0^2}{n^2} \cdot \frac{qrs^2}{qe^s + r + we^{-s}} = \frac{\theta_0^2}{n^2} \cdot \frac{qrs^2e^s}{qe^{2s} + re^s + w}. \tag{77}$$

$$\begin{aligned}
 I_2 &= rw\Delta^2 \int_{\mathcal{C}(\theta_0)} \frac{\theta_0^{-n}(\theta_0 + \Delta)^{-n} dx}{q(\theta_0 - \Delta)^{-n} \mathcal{I}\{x \in \mathcal{C}(\theta_0 - \Delta)\} + r\theta_0^{-n} + w(\theta_0 + \Delta)^{-n}} \\
 &= rw\Delta^2 \cdot \frac{(\theta_0 - \Delta)^n \theta_0^{-n} (\theta_0 + \Delta)^{-n}}{q(\theta_0 - \Delta)^{-n} + r\theta_0^{-n} + w(\theta_0 + \Delta)^{-n}} + \\
 &\quad rw\Delta^2 \cdot \frac{[\theta_0^n - (\theta_0 - \Delta)^n] \theta_0^{-n} (\theta_0 + \Delta)^{-n}}{r\theta_0^{-n} + w(\theta_0 + \Delta)^{-n}} \\
 &= rw\Delta^2 \cdot \frac{(1 - \Delta/\theta_0)^n}{q[(1 + \Delta/\theta_0)/(1 - \Delta/\theta_0)]^{-n} + r(1 + \Delta/\theta_0)^n + w} + \\
 &\quad rw\Delta^2 \cdot \frac{[1 - (1 - \Delta/\theta_0)^n]}{r(1 + \Delta/\theta_0)^n + w} \\
 &\sim \frac{\theta_0^2}{n^2} \cdot rws^2 \left( \frac{e^{-s}}{qe^{2s} + re^s + w} + \frac{1 - e^{-s}}{re^s + w} \right).
 \end{aligned} \tag{78}$$

Similarly,

$$I_3 \sim \frac{\theta_0^2}{n^2} \cdot \frac{4qws^2}{qe^{2s} + re^s + w}. \tag{79}$$

Thus,

$$\begin{aligned}
 r(\theta_0) &\geq \theta_0^2 \cdot \sup_{s \geq 0} \sup_{(q,r,w) \in \mathbb{R}_+^3: q+r+w=1} s^2 \cdot \left[ \frac{qre^s + 4qw + rwe^{-s}}{qe^{2s} + re^s + w} + \frac{rw(1 - e^{-s})}{re^s + w} \right] \\
 &\approx 0.4624\theta_0^2,
 \end{aligned} \tag{80}$$

w.r.t.  $\zeta_n = n^2$ , which is a significant improvement of nearly 50% over the previous bound,  $0.3102\theta_0^2$ , which was obtained on the basis of two test points, let alone the bound of Theorem 1, which was  $0.2414\theta_0^2$ . This concludes Example 8.

In general, the integrals  $I_1$ ,  $I_2$ , and  $I_3$  are not easy to calculate due to the summations at the denominators of the integrands. One way to proceed is to apply Lemma 1 to the sum of  $k = 3$  terms, but this would introduce two additional parameters to be optimized. Returning to the earlier shorthand notation of  $a$ ,  $b$ , and  $c$ , a different approach to get rid of summations at the denominators, at the expense of some loss of tightness, is the following:

$$\begin{aligned}
 \frac{ab + bc + 4ac}{a + b + c} &= \frac{a(b + 2c)}{a + b + c} + \frac{c(b + 2a)}{a + b + c} \\
 &\geq \frac{a(b + 2c)}{a + b + 2c} + \frac{c(b + 2a)}{2a + b + c} \\
 &\geq \frac{a(b + 2c)}{2 \max\{a, b + 2c\}} + \frac{c(b + 2a)}{2 \max\{2a + b, c\}}
 \end{aligned} \tag{81}$$

$$\begin{aligned}
 &= \frac{1}{2} \cdot (\min\{a, b + 2c\} + \min\{2a + b, c\}) \\
 &\geq \frac{1}{2} \cdot (\min\{a, b\} + \min\{b, c\}) \\
 &= \frac{1}{2} \cdot [\min\{qp(x|\theta_0 - \Delta), rp(x|\theta_0)\} + \min\{rp(x|\theta_0), wp(x|\theta_0 + \Delta)\}],
 \end{aligned}$$

and so,

$$\begin{aligned}
 R_n(\Theta) &\geq \frac{\Delta^2}{2} \left[ \int_{\mathbb{R}^n} \min\{qp(x|\theta_0 - \Delta), rp(x|\theta_0)\} dx + \right. \\
 &\quad \left. \int_{\mathbb{R}^n} \min\{rp(x|\theta_0), wp(x|\theta_0 + \Delta)\} dx \right] \\
 &= \frac{\Delta^2}{2} \left[ (q + r)P_e\left(\frac{q}{q+r}, \theta_0 - \Delta, \theta_0\right) + \right. \\
 &\quad \left. (r + w)P_e\left(\frac{r}{r+w}, \theta_0, \theta_0 + \Delta\right) \right].
 \end{aligned} \tag{82}$$

Note that by setting  $w = 0$ , we recover the bound obtained by integrating (53), and therefore, by optimizing  $w$ , the resulting bound cannot be worse. A slightly different (better, but more complicated) route in (81) is to apply Lemma 1 with  $k = 2$  in the following manner:

$$\begin{aligned}
 \frac{ab + bc + 4ac}{a + b + c} &= \frac{a(b + 2c)}{a + b + c} + \frac{c(b + 2a)}{a + b + c} \\
 &\geq \frac{a(b + 2c)}{a + b + 2c} + \frac{c(b + 2a)}{2a + b + c} \\
 &= \frac{a(b + 2c)}{\min_u \max\{a/u, (b + 2c)/(1 - u)\}} + \frac{c(b + 2a)}{\min_v \max\{(2a + b)/(1 - v), c/v\}} \\
 &= \max_u \min\{(1 - u)a, u(b + 2c)\} + \max_v \min\{v(2a + b), (1 - v)c\} \\
 &\geq \max_u \min\{(1 - u)a, ub\} + \max_v \min\{vb, (1 - v)c\} \\
 &= \max_u \min\{(1 - u)qp(x|\theta_0 - \Delta), urp(x|\theta_0)\} + \\
 &\quad \max_v \min\{v rp(x|\theta_0), (1 - v)wp(x|\theta_0 + \Delta)\},
 \end{aligned} \tag{83}$$

and so,

$$\begin{aligned}
 R_n(\Theta) &\geq \Delta^2 \cdot \left[ \max_u \int_{\mathbb{R}^n} \min\{(1 - u)qp(x|\theta_0 - \Delta), urp(x|\theta_0)\} dx + \right. \\
 &\quad \left. \max_v \int_{\mathbb{R}^n} \min\{v rp(x|\theta_0), (1 - v)wp(x|\theta_0 + \Delta)\} dx \right] \\
 &= \Delta^2 \cdot \left\{ \max_u [(1 - u)q + ur] \cdot P_e\left(\frac{(1 - u)q}{(1 - u)q + ur}, \theta_0 - \Delta, \theta_0\right) + \right. \\
 &\quad \left. \max_v [vr + (1 - v)w] \cdot P_e\left(\frac{vr}{vr + (1 - v)w}, \theta_0, \theta_0 + \Delta\right) \right\}.
 \end{aligned} \tag{84}$$

We have just proved the following result:

**Theorem 3.** For the MSE case,

$$\begin{aligned}
 R_n(\Theta) &\geq \sup_{\theta_0, \Delta, q, r, w, (u, v) \in [0, 1]^2} \Delta^2 \cdot \left\{ \max_u [(1 - u)q + ur] \cdot P_e\left(\frac{(1 - u)q}{(1 - u)q + ur}, \theta_0 - \Delta, \theta_0\right) + \right. \\
 &\quad \left. \max_v [vr + (1 - v)w] \cdot P_e\left(\frac{vr}{vr + (1 - v)w}, \theta_0, \theta_0 + \Delta\right) \right\}.
 \end{aligned} \tag{85}$$

where the maximum over  $q, r, w$  is under the constraints that they are all non-negative and sum to unity.

**Example 9.** Revisiting Example 4, it is natural that both error probabilities would correspond to a prior of  $1/2$ . This dictates the relations,

$$u = \frac{q}{q+r} \tag{86}$$

$$v = \frac{w}{w+r}, \tag{87}$$

and so, the bound becomes

$$\begin{aligned} R_n(\Theta) &\geq \max_{\Delta, q, w, r} \Delta^2 \left( \frac{2qr}{q+r} + \frac{2wr}{w+r} \right) \cdot Q\left(\frac{\sqrt{n}\Delta}{2\sigma}\right) \\ &= \left[ \max_{s \geq 0} \left( \frac{2s\sigma}{\sqrt{n}} \right)^2 Q(s) \right] \cdot \max_{\{(q,r): q \geq 0, r \geq 0, q+r \leq 1\}} \left\{ \frac{2qr}{q+r} + \frac{2r(1-q-r)}{1-q} \right\} \\ &= \frac{\sigma^2}{n} \cdot \max_{s \geq 0} \{4s^2 Q(s)\} \cdot 0.6862 \\ &= \frac{\sigma^2}{n} \cdot 0.6629 \cdot 0.6862 \\ &\approx \frac{0.4549\sigma^2}{n}, \end{aligned} \tag{88}$$

which is an improvement on the bound of Example 4, which was  $0.3314\sigma^2/n$ . Similarly, in Example 6, for smooth signals, the multiplicative constant also improves from 0.3314 to 0.4549, and for the rectangular pulse, it improves from 0.7544 to 1.0352 (about 37% improvement in all of them). This concludes Example 9.

**Example 10.** Revisiting Example 2, we now have

$$\begin{aligned} R_n(\Theta) &\geq \Delta^2 \cdot \left[ \max_u \min\{(1-u)q, ur\alpha\} + \max_v \min\{vr, (1-v)w\alpha\} \right] \\ &= \Delta^2 \left( \frac{qr\alpha}{q+r\alpha} + \frac{wr\alpha}{r+w\alpha} \right) \\ &= \Delta^2 \left( \frac{qr\alpha}{q+r\alpha} + \frac{r(1-q-r)\alpha}{r+(1-q-r)\alpha} \right) \\ &= \frac{s^2}{n^2} e^{-s} r \left( \frac{q}{q+re^{-s}} + \frac{1-q-r}{r+(1-q-r)e^{-s}} \right) \\ &= \frac{s^2}{n^2} r \left( \frac{q}{qe^s+r} + \frac{1-q-r}{re^s+(1-q-r)} \right), \end{aligned} \tag{89}$$

whose maximum is  $0.3909/n^2$ , an improvement relative to the earlier bound of  $0.3102/n^2$ . This concludes Example 10.

### 5. Extensions to the Vector Case

In this section, we outline extensions of some of our findings in Sections 3 and 4 to the vector case. Let  $\theta$  be a parameter vector of dimension  $d$  and  $\Theta \subseteq \mathbb{R}^d$ . Let  $\rho(\varepsilon)$  be a convex loss function that depends on the  $d$ -dimensional error vector  $\varepsilon$  only via its norm,  $\|\varepsilon\|$  (that is,  $\rho$  has radial symmetry).

First, observe that Theorem 1 extends verbatim to the vector case, as nothing in the proof of Theorem 1 is based on any assumption or property that holds only if  $\theta$  is a scalar parameter.

Corollary 1 can also be extended by letting the two test points of the auxiliary hypothesis testing problem,  $\theta_0$  and  $\theta_1$ , be selected such that the distance between them,  $\|\theta_1 - \theta_0\|$ ,

would decay at an appropriate rate as a function of  $n$ , so as to make the probability of error,  $\max_q P_e(q, \theta_0, \theta_1)$ , converge to a positive constant as  $n \rightarrow \infty$ , as was achieved in Corollary 1 for the scalar case. But in the vector case considered now, there is an additional degree of freedom, which is the *direction* of the displacement vector,  $\theta_1 - \theta_0$ . This direction can now be optimized so as to yield the largest (hence the tightest) lower bound. To be specific, in the vector case, Corollary 1 remains the same except that now  $s$  should be thought of as a  $d$ -dimensional vector rather than a scalar,  $\rho(u)$  in the denominator of (19) should be replaced by  $\rho(v \cdot u)$ , where  $v \in \mathbb{R}^d$  is an arbitrary fixed non-zero vector (for example, any unit norm vector in the case of MSE), and the supremum in (20) should be taken over  $\mathbb{R}^d$ . To demonstrate this point, we now revisit Example 5, but this time, for the vector case.

**Example 11.** Consider the case where  $p(x|\theta) = \prod_{i=1}^n p(x_i|\theta)$ , with each factor in this product PDF being given by a  $d$ -dimensional exponential family,

$$p(x|\theta) = \frac{\exp\{\theta' T(x)\}}{Z(\theta)} \tag{90}$$

where  $\theta' T(x)$  is the inner product of the  $d$ -dimensional parameter vector  $\theta$  and a  $d$ -dimensional vector of statistics,  $T(x) = (T_1(x), T_2(x), \dots, T_d(x))$ , and

$$Z(\theta) = \int_{\mathbb{R}} \exp\{\theta' T(x)\} dx, \tag{91}$$

provided that the integral converges. In the above notation, both  $\theta$  and  $T(x)$  are understood to be column vectors and  $\theta'$  denotes transposition of  $\theta$  to a row vector. Similarly as in Corollary 1, to obtain a local bound at a given  $\theta$ , we let  $\theta_0 = \theta$  and  $\theta_1 = \theta + \frac{2s}{\sqrt{n}}$ , where  $s \in \mathbb{R}^d$ . A simple extension of the derivation in Example 5 (using the CLT) yields

$$P_e^\infty(\theta, s) = Q\left(\sqrt{s' I(\theta) s}\right), \tag{92}$$

where  $s$  is considered a column vector, the superscript prime denotes vector transposition as before, and  $I(\theta) = \nabla^2 \ln Z(\theta)$  is the  $d \times d$  Fisher information matrix of the exponential family. Considering the case of the MSE,  $\rho(\epsilon) = \|\epsilon\|^2$  with  $v$  being any unit norm vector, we have

$$w(s) = \lim_{u \rightarrow 0} \frac{\|s \cdot u\|^2}{\|v \cdot u\|^2} = \|s\|^2, \tag{93}$$

and then the following lower bound is obtained w.r.t.  $1/\|v/\sqrt{n}\|^2 = n$ :

$$\begin{aligned} r(\theta) &\geq \sup_{s \in \mathbb{R}^d} \left\{ 2\|s\|^2 Q\left(\sqrt{s' I(\theta) s}\right) \right\} \\ &= \sup_{t \geq 0} \left\{ 2t^2 \cdot \max_{\{s: \|s\|^2=t^2\}} Q\left(\sqrt{s' I(\theta) s}\right) \right\} \\ &= \sup_{t \geq 0} \left\{ 2t^2 \cdot Q\left(t\sqrt{\lambda_{\min}(\theta)}\right) \right\} \\ &= \frac{1}{\lambda_{\min}(\theta)} \sup_{\tau \geq 0} \left\{ 2\tau^2 \cdot Q(\tau) \right\} \\ &\approx \frac{0.3314}{\lambda_{\min}(\theta)}, \end{aligned} \tag{94}$$

where  $\lambda_{\min}(\theta)$  is the smallest eigenvalue of  $I(\theta)$ . In this example, it is apparent that the chosen direction of the vector  $s$  is that of the eigenvector corresponding to the smallest eigenvalue of the Fisher information matrix. This completes Example 11.

In the vector case considered in this section, it is also instructive to extend the scope from two test points,  $\theta_0$  and  $\theta_1$ , to multiple test points,  $\theta_0, \theta_1, \dots, \theta_{m-1}$ , along with the corresponding weights (or priors),  $q_0, q_1, \dots, q_{m-1}$  (see the exposition at the end of Section 2). Interestingly, this will also lead to new bounds even for the case of  $m = 2$  test points.

To this end, we also consider a set of  $m$  unitary transformation matrices,  $T_0, T_1, T_2, \dots, T_{m-1}$ , with the properties: (i)  $\theta \in \Theta$  if and only if  $T_i\theta \in \Theta$  for all  $i = 0, 1, \dots, m - 1$ , and (ii)  $T_0 + T_1 + \dots + T_{m-1} = 0$ . For example, if  $d = 2$ , take  $T_i$  to be matrices of rotation by  $2\pi i/m$ ,  $i = 0, 1, \dots, m - 1$ .

**Theorem 4.** Let  $\theta_0, \theta_1, \dots, \theta_{m-1} \in \Theta \subseteq \mathbb{R}^d$ ,  $d \geq 1$  and  $q_0, q_1, \dots, q_{m-1}$  be given, and let  $T_0, T_1, \dots, T_{m-1}$  be unitary transformations that sum to zero, as described in the above paragraph. Then, for a convex loss function  $\rho(\varepsilon)$  that depends on  $\varepsilon$  only via  $\|\varepsilon\|$ , we have:

$$R_n(\Theta) \geq m \cdot \rho\left(\frac{1}{m} \sum_{i=0}^{m-1} T_i\theta_i\right) \cdot \int_{\mathbb{R}^n} \min\{q_0 \cdot p(\mathbf{x}|\theta_0), \dots, q_{m-1}p(\mathbf{x}|\theta_{m-1})\}dx. \tag{95}$$

As explained in Section 2, the integral on the right-hand side can be interpreted as the probability of list error with a list size of  $m - 1$ .

**Proof of Theorem 4.** The proof is a direct extension of the proof of Theorem 1:

$$\begin{aligned} \sup_{\theta \in \Theta} \mathbb{E}_{\theta} \{\rho(g_n[\mathbf{X}] - \theta)\} &\geq \sum_{i=0}^{m-1} q_i \mathbb{E}_{\theta_i} \{\rho(g_n[\mathbf{X}] - \theta_i)\} \\ &= \int_{\mathbb{R}^n} \left[ \sum_{i=0}^{m-1} q_i \cdot p(\mathbf{x}|\theta_i) \rho(g_n[\mathbf{x}] - \theta_i) \right] dx \\ &\geq m \cdot \int_{\mathbb{R}^n} \min\{q_0 \cdot p(\mathbf{x}|\theta_0), \dots, q_{m-1}p(\mathbf{x}|\theta_{m-1})\} \times \\ &\quad \left[ \frac{1}{m} \sum_{i=0}^{m-1} \rho(\theta_i - g_n[\mathbf{X}]) \right] dx \\ &\stackrel{(a)}{=} m \cdot \int_{\mathbb{R}^n} \min\{q_0 \cdot p(\mathbf{x}|\theta_0), \dots, q_{m-1}p(\mathbf{x}|\theta_{m-1})\} \times \\ &\quad \left[ \frac{1}{m} \sum_{i=0}^{m-1} \rho(T_i(\theta_i - g_n[\mathbf{X}])) \right] dx \\ &\stackrel{(b)}{\geq} m \cdot \int_{\mathbb{R}^n} \min\{q_0 \cdot p(\mathbf{x}|\theta_0), \dots, q_{m-1}p(\mathbf{x}|\theta_{m-1})\} \times \\ &\quad \rho\left(\frac{1}{m} \sum_{i=0}^{m-1} T_i(\theta_i - g_n[\mathbf{X}])\right) dx \\ &\stackrel{(c)}{=} m \cdot \int_{\mathbb{R}^n} \min\{q_0 \cdot p(\mathbf{x}|\theta_0), \dots, q_{m-1}p(\mathbf{x}|\theta_{m-1})\} \times \\ &\quad \rho\left(\frac{1}{m} \sum_{i=0}^{m-1} T_i\theta_i\right) dx \\ &= m\rho\left(\frac{1}{m} \sum_{i=0}^{m-1} T_i\theta_i\right) \cdot \int_{\mathbb{R}^n} \min\{q_0 \cdot p(\mathbf{x}|\theta_0), \dots, q_{m-1}p(\mathbf{x}|\theta_{m-1})\}dx, \end{aligned}$$

where in (a), we have used the unitarity (and hence the norm-preserving property) of  $\{T_i\}$ , as  $\rho(\cdot)$  is assumed to depend only on the norm of the error vector; in (b), we used the convexity of  $\rho$ , and in (c), we used the fact that  $\sum_{i=0}^{m-1} T_i = 0$ , which implies that  $\sum_{i=0}^{m-1} T_i g_n[\mathbf{X}] = 0$ , thus making the bound independent of the estimator,  $g_n$ . This completes the proof of Theorem 4.  $\square$

Theorem 1 is a special case where  $m = 2$ ,  $T_0 = I$ , and  $T_1 = -I$ , where  $I$  is the  $d \times d$  identity matrix. The integral associated with the lower bound of Theorem 4 might

not be trivial to evaluate in general for  $m \geq 3$ . However, there are some choices of the auxiliary parameters that may facilitate calculations. One such choice is as follows. For some positive integer  $k < m$ , take  $\theta_0 = \theta_1 = \dots = \theta_{k-1} \triangleq \vartheta_0$ , for some  $\vartheta_0 \in \Theta$ ,  $q_0 = q_1 = \dots = q_{k-1} \triangleq Q/k$ , for some  $Q \in (0, 1)$ ,  $\theta_k = \theta_{k+1} = \dots = \theta_{m-1} \triangleq \vartheta_1$ , for some  $\vartheta_1 \in \Theta$ , and finally,  $q_k = q_{k+1} = \dots = q_{m-1} = (1 - Q)/(m - k)$ . The integrand then becomes the minimum between two functions only, as in Section 3. Denoting  $\alpha = k/m$ , the bound then becomes

$$\begin{aligned} R_n(\Theta) &\geq m\rho\left(\frac{1}{m}\sum_{i=0}^{k-1}T_i(\vartheta_0 - \vartheta_1)\right) \cdot \int_{\mathbb{R}^n} \min\left\{\frac{Q}{k} \cdot p(x|\vartheta_0), \frac{1-Q}{m-k} \cdot p(x|\vartheta_1)\right\} dx \\ &= \rho\left(\frac{1}{m}\sum_{i=0}^{k-1}T_i(\vartheta_0 - \vartheta_1)\right) \cdot \int_{\mathbb{R}^n} \min\left\{\frac{Q}{\alpha} \cdot p(x|\vartheta_0), \frac{1-Q}{1-\alpha} \cdot p(x|\vartheta_1)\right\} dx \\ &= \rho\left(\frac{1}{m}\sum_{i=0}^{k-1}T_i(\vartheta_0 - \vartheta_1)\right) \cdot \left(\frac{Q}{\alpha} + \frac{1-Q}{1-\alpha}\right) \cdot P_e\left(\frac{(1-\alpha)Q}{(1-\alpha)Q + \alpha(1-Q)}, \vartheta_0, \vartheta_1\right). \end{aligned} \tag{96}$$

Redefining

$$q = \frac{(1-\alpha)Q}{(1-\alpha)Q + \alpha(1-Q)}, \tag{97}$$

we have

$$\frac{Q}{\alpha} + \frac{1-Q}{1-\alpha} = \frac{1}{1-\alpha-q+2\alpha q}, \tag{98}$$

and the following corollary to Theorem 2 is obtained.

**Corollary 2.** *Let the conditions of Theorem 4 be satisfied. Then,*

$$\begin{aligned} R_n(\Theta) &\geq \sup_{\vartheta_0, \vartheta_1, \alpha, q} \rho\left(\frac{1}{m}\sum_{i=0}^{k-1}T_i(\vartheta_0 - \vartheta_1)\right) \cdot \frac{P_e(q, \vartheta_0, \vartheta_1)}{1-\alpha-q+2\alpha q} \\ &= \sup_{\vartheta_0, \vartheta_1, q} \rho\left(\frac{1}{m}\sum_{i=0}^{k-1}T_i(\vartheta_0 - \vartheta_1)\right) \cdot \frac{P_e(q, \vartheta_0, \vartheta_1)}{\min\{q, 1-q\}}. \end{aligned} \tag{99}$$

Note that if  $m$  is even,  $\alpha = \frac{1}{2}$ , and  $T_0 = I$ ; then, we are actually back to the bound of  $m = 2$ , and so, the optimal bound for even  $m > 2$  cannot be worse than our bound of  $m = 2$ . We do not have, however, precisely the same argument for an odd  $m$ , but for a large  $m$ , it becomes immaterial if  $m$  is even or odd. In its general form, the bound of Theorem 4 is a heavy optimization problem, as we have the freedom to optimize  $\theta_0, \dots, \theta_{m-1}, T_0, \dots, T_{m-1}$  (under the constraints that they are all unitary and sum to zero), and  $q_0, \dots, q_{m-1}$  (under the constraints that they are all non-negative and sum to unity).

Another relatively convenient choice is to take  $\theta_i = T_i^{-1}\theta_0, i = 1, \dots, m - 1$  to obtain another corollary to Theorem 4:

**Corollary 3.** *Let the conditions of Theorem 2 be satisfied. Then,*

$$R_n(\Theta) \geq \sup_{\theta_0, T_0, \dots, T_{m-1}, q_0, \dots, q_{m-1}} m\rho(\theta_0) \cdot \int_{\mathbb{R}^n} \min\{q_0 p(x|\theta_0), q_1 p(x|T_1^{-1}\theta_0), \dots, q_{m-1} p(x|T_{m-1}^{-1}\theta_0)\} dx. \tag{100}$$

**Example 12.** *To demonstrate a calculation of the extended lower bound for  $m = 3$ , consider the following model. We are observing a noisy signal,*

$$Z_i = \vartheta\phi_i + (\vartheta + \zeta)\psi_i + N_i, \quad i = 1, 2, \dots, n, \tag{101}$$

where  $\vartheta$  is the desired parameter to be estimated,  $\zeta$  is a nuisance parameter, taking values within an interval  $[-\delta, \delta]$  for some  $\delta > 0$ ,  $\{N_i\}$  are i.i.d. Gaussian random variables with zero mean and

variance  $\sigma^2$ , and  $\phi_i$  and  $\psi_i$  are two given orthogonal waveforms with  $\sum_{i=1}^n \phi_i^2 = \sum_{i=1}^n \psi_i^2 = n$ . Suppose we are interested in estimating  $\vartheta$  based on the sufficient statistics  $X = \frac{1}{n} \sum_{i=1}^n Z_i \phi_i$  and  $Y = \frac{1}{n} \sum_{i=1}^n Z_i \psi_i$ , which are jointly Gaussian random variables with the mean vector  $(\vartheta, \vartheta + \zeta)$  and the covariance matrix  $\frac{\sigma^2}{n} \cdot I$ ,  $I$  being the  $2 \times 2$  identity matrix. We denote realizations of  $(X, Y)$  by  $(x, y)$ . Let us also denote  $\theta = (\vartheta, \vartheta + \zeta)$ . Since we are interested only in estimating  $\vartheta$ , our loss function will depend only on the estimation error of the first component of  $\theta$ , which is  $\vartheta$ . Consider the choice  $m = 3$  and let  $T_i$  be counter-clockwise rotation transformations by  $2\pi i/3$ ,  $i = 0, 1, 2$ . For a given  $\Delta \in (0, \delta]$ , let us select  $\theta_0 = (-\Delta, 0)$ ,  $\theta_1 = T_1^{-1}\theta_0 = (\Delta/2, \Delta\sqrt{3}/2)$ , and  $\theta_2 = T_2^{-1}\theta_0 = (\Delta/2, -\Delta\sqrt{3}/2)$ . Finally, let  $q_0 = q_1 = q_2 = \frac{1}{3}$ . In order to calculate the integral

$$I = \int_{\mathbb{R}^2} \min \left\{ \frac{1}{3} p(x, y | \theta_0), \frac{1}{3} p(x, y | \theta_1), \frac{1}{3} p(x, y | \theta_2) \right\} dx dy,$$

the plane  $\mathbb{R}^2$  can be partitioned into three slices over which the integrals contributed are equal. In each such region, the smallest  $p(x, y | \theta_i)$  is integrated. In other words, every  $p(x, y | \theta_i)$  in its turn is integrated over the region whose Euclidean distance to  $\theta_i$  is larger than the distances to the other two values of  $\theta$ . For  $\theta_0 = (-\Delta, 0)$ , this is the region  $\{(x, y) : x \geq 0, |y| \leq x\sqrt{3}\}$ . The factor of  $\frac{1}{3}$  cancels with the three identical contributions from  $\theta_0, \theta_1$ , and  $\theta_2$  due to the symmetry. Therefore,

$$\begin{aligned} I &= \int_0^\infty dx \int_{-x\sqrt{3}}^{x\sqrt{3}} dy \cdot p(x, y | \theta_0) \\ &= \int_0^\infty dx \int_{-x\sqrt{3}}^{x\sqrt{3}} \frac{n}{2\pi\sigma^2} \exp \left\{ -\frac{n(x + \Delta)^2 + ny^2}{2\sigma^2} \right\} dy \\ &= \sqrt{\frac{n}{2\pi\sigma^2}} \cdot \int_0^\infty \exp \left\{ -\frac{n(x + \Delta)^2}{2\sigma^2} \right\} \left[ 1 - 2Q \left( \frac{x\sqrt{3n}}{\sigma} \right) \right] dx. \end{aligned} \tag{102}$$

Our next mathematical manipulations in this example are in the spirit of the passage from Theorem 1 to Corollary 1, that is, selecting the test points increasingly close to each other as functions of  $n$ , so that the probability of list error would tend to a positive constant. To this end, we change the integration variable  $x$  to  $u = x\sqrt{n}/\sigma$  and select  $\Delta = s\sigma/\sqrt{n}$  for some  $s \geq 0$  to be optimized later. Then,

$$\begin{aligned} I &= \sqrt{\frac{n}{2\pi\sigma^2}} \cdot \int_0^\infty \exp \left\{ -\frac{n(u\sigma/\sqrt{n} + s\sigma/\sqrt{n})^2}{2\sigma^2} \right\} \left[ 1 - 2Q(u\sqrt{3}) \right] d \left( \frac{u\sigma}{\sqrt{n}} \right) \\ &= \frac{1}{\sqrt{2\pi}} \cdot \int_0^\infty e^{-(u+s)^2/2} \cdot [1 - 2Q(u\sqrt{3})] du. \end{aligned} \tag{103}$$

The MSE bound then becomes

$$\begin{aligned} R_n(\theta, g_n) &\geq \sup_{s \geq 0} \left\{ 3 \cdot \left( \frac{s\sigma}{\sqrt{n}} \right)^2 \cdot \frac{1}{\sqrt{2\pi}} \cdot \int_0^\infty e^{-(u+s)^2/2} \cdot [1 - 2Q(u\sqrt{3})] du \right\} \\ &= \frac{\sigma^2}{n} \cdot \sup_{s \geq 0} \left\{ \frac{3s^2}{\sqrt{2\pi}} \cdot \int_0^\infty e^{-(u+s)^2/2} \cdot [1 - 2Q(u\sqrt{3})] du \right\} \\ &\approx \frac{0.2514\sigma^2}{n}. \end{aligned} \tag{104}$$

This bound is not as tight as the corresponding bound of  $m = 2$ , which results in  $0.3314\sigma^2/n$ , but it should be kept in mind that here, we have not attempted to optimize the choices of  $\theta_0, \theta_1, \theta_2, T_0, T_1, T_2, q_0$ , and  $q_1$ . Instead, we have chosen these parameter values from considerations of computational convenience, just to demonstrate the calculation. This concludes Example 12.

Finally, a comment is in order regarding the possible extensions of Section 4 to the vector case. Such extensions are conceptually straightforward whenever the loss function of the error vector is given by the sum of losses associated with the different components of the



error vector. Most notably, when the loss function is the MSE,  $\rho(\epsilon) = \|\epsilon\|^2$ , each component of the estimation error can be handled separately by the methods of Section 4. Of course, the two or three test points should be chosen such that they differ in all components of the parameter vector. In the case of three test points, it makes sense to select them equally spaced along a certain straight line of a general direction in  $\Theta$ . We will not pursue any further this extension in the framework of this work.

**Funding:** This research received no external funding.

**Data Availability Statement:** The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The author declares no conflict of interest.

### Appendix A

**Proof of Equation (44).** Consider the Taylor series expansion,

$$s(t, \theta + \delta) = s(t, \theta) + \delta \cdot \dot{s}(t, \theta) + \frac{\delta^2}{2} \cdot \ddot{s}(t, \theta) + o(\delta^2), \tag{A1}$$

where  $\dot{s}(t, \theta)$  and  $\ddot{s}(t, \theta)$  are the first two partial derivatives of  $s(t, \theta)$  w.r.t.  $\theta$ . Correlating both sides with  $s(t, \theta)$  yields

$$\int_0^T s(t, \theta)s(t, \theta + \delta)dt = E + \delta \cdot \int_0^T s(t, \theta)\dot{s}(t, \theta)dt + \frac{\delta^2}{2} \cdot \int_0^T s(t, \theta)\ddot{s}(t, \theta)dt + o(\delta^2). \tag{A2}$$

Now,

$$\int_0^T s(t, \theta)\dot{s}(t, \theta)dt = \frac{1}{2} \cdot \frac{\partial}{\partial \theta} \left\{ \int_0^T s^2(t, \theta)dt \right\} = \frac{1}{2} \cdot \frac{\partial E}{\partial \theta} = 0, \tag{A3}$$

since the energy  $E$  is assumed independent of  $\theta$ . Also, since  $\frac{\partial^2 E}{\partial \theta^2} = 0$ , we have

$$\begin{aligned} 0 &= \frac{\partial^2}{\partial \theta^2} \left\{ \int_0^T s^2(t, \theta)dt \right\} \\ &= \frac{\partial}{\partial \theta} \left\{ 2 \cdot \int_0^T s(t, \theta)\dot{s}(t, \theta)dt \right\} \\ &= 2 \cdot \int_0^T \dot{s}^2(t, \theta)dt + 2 \cdot \int_0^T s(t, \theta)\ddot{s}(t, \theta)dt, \end{aligned} \tag{A4}$$

which yields

$$\int_0^T s(t, \theta)\ddot{s}(t, \theta)dt = - \int_0^T \dot{s}^2(t, \theta)dt, \tag{A5}$$

and so,

$$\int_0^T s(t, \theta)s(t, \theta + \delta)dt = E - \frac{\delta^2}{2} \int_0^T \dot{s}^2(t, \theta)dt + o(\delta^2), \tag{A6}$$

and hence,

$$\varrho(\theta, \theta + \Delta) = 1 - \frac{\delta^2}{2E} \int_0^T \dot{s}^2(t, \theta)dt + o(\delta^2). \tag{A7}$$

□

### Appendix B

**Proof of Lemma 1.** First, observe that

$$\sum_{i=1}^k a_i = \sum_{i=1}^k r_i \cdot \frac{a_i}{r_i} \leq \max \left\{ \frac{a_1}{r_1}, \dots, \frac{a_k}{r_k} \right\}, \tag{A8}$$

and since the inequality  $\sum_{i=1}^k a_i \leq \max\{a_1/r_1, \dots, a_k/r_k\}$  applies to all  $\mathbf{r} \in \mathcal{S}$ , it also applies to the infimum of  $\max\{a_1/r_1, \dots, a_k/r_k\}$  over  $\mathcal{S}$ , thus establishing the inequality “ $\leq$ ” between the two sides. To establish the “ $\geq$ ” inequality,  $\mathbf{r}^* \in \mathcal{S}$  is defined to be the vector whose components are given by  $r_i^* = a_i / \sum_{j=1}^k a_j$ . Then,

$$\inf_{(r_1, \dots, r_k) \in \mathcal{S}} \max\left\{\frac{a_1}{r_1}, \dots, \frac{a_k}{r_k}\right\} \leq \max\left\{\frac{a_1}{r_1^*}, \dots, \frac{a_k}{r_k^*}\right\} = \sum_{i=1}^k a_i. \quad (\text{A9})$$

This completes the proof of Lemma 1.  $\square$

## References

1. Van Trees, H.L. *Detection, Estimation and Modulation Theory: Part I*; John Wiley & Sons, Inc.: New York, NY, USA, 1968.
2. Bobrovsky, B.; Zakai, M. A lower bound on the estimation error for certain diffusion processes. *IEEE Trans. Inform. Theory* **1976**, *22*, 45–52. [\[CrossRef\]](#)
3. Bellini, S.; Tartara, G. Bounds on error in signal parameter estimation. *IEEE Trans. Commun.* **1974**, *22*, 340–342. [\[CrossRef\]](#)
4. Chazan, D.; Zakai, M.; Ziv, J. Improved lower bounds on signal parameter estimation. *IEEE Trans. Inform. Theory* **1975**, *21*, 90–93. [\[CrossRef\]](#)
5. Weiss, A.J. *Fundamental Bounds in Parameter Estimation*. Ph.D. Thesis, Department of Electrical Engineering—Systems, Tel Aviv University, Tel Aviv, Israel, June 1985.
6. Weiss, A.J.; Weinstein, E. A lower bound on the mean square error in random parameter estimation. *IEEE Trans. Inf. Theory* **1985**, *31*, 680–682. [\[CrossRef\]](#)
7. Van Trees, H.L.; Bell, K.L. (Eds.) *Bayesian Bounds for Parametric Estimation and Nonlinear Filtering/Tracking*; John Wiley & Sons, Inc.: New York, NY, USA, 2007.
8. Fisher, R.A. On the mathematical foundations of theoretical statistics. *Philos. Trans. R. Soc.* **1922**, *222*, 309.
9. Dugue, D. Application des proprietes de la limite au sens du calcul des probabilites a l’etudes des diverses questions d’estimation. *Ecol. Poly.* **1937**, *3*, 305–372.
10. Frechet, M. Sur l’extension de certaines evaluations statistiques au cas de petits echantillons. *Rev. Inst. Int. Stat.* **1943**, *11*, 182–205. [\[CrossRef\]](#)
11. Rao, C.R. Information and accuracy attainable in the estimation of statistical parameters. *Bull. Calcutta Math. Soc.* **1945**, *37*, 81–91.
12. Cramér, H. *Mathematical Methods of Statistics*; Princeton University Press: Princeton, NJ, USA, 1946.
13. Bhattacharyya, A. On some analogues of the amount of information and their use in statistical estimation. *Sankhya India J. Stat.* **1946**, *8*, 1–14.
14. Barankin, E.W. Locally best unbiased estimators. *Ann. Math. Stat.* **1949**, *20*, 477–501. [\[CrossRef\]](#)
15. Chapman, D.G.; Robbins, H. Minimum variance estimation without regularity assumption. *Ann. Math. Stat.* **1951**, *22*, 581–586. [\[CrossRef\]](#)
16. Fraser, D.A.; Guttman, I. Bhattacharyya bound without regularity assumptions. *Ann. Math. Stat.* **1952**, *23*, 629–632. [\[CrossRef\]](#)
17. Kiefer, J. On minimum variance estimation. *Ann. Math. Stat.* **1952**, *23*, 627–629. [\[CrossRef\]](#)
18. Ziv, J.; Zakai, M. Some lower bounds on signal parameter estimation. *IEEE Trans. Inform. Theory* **1969**, *15*, 386–391. [\[CrossRef\]](#)
19. Hájek, J. Local asymptotic minimax and admissibility in estimation. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*; University of California Press: Berkeley, CA, USA, 1972; pp. 175–194.
20. Le Cam, L.M. Convergence of estimates under dimensionality restrictions. *Ann. Stat.* **1973**, *1*, 38–53.
21. Assouad, P. Deux remarques sur l’estimation. *C. R. L’Acad. Sci. Paris Ser. Math.* **1983**, *296*, 1021–1024.
22. Fano, R.M. *Transmission of Information: A Statistical Theory of Communication*; The M.I.T. Press: Cambridge, MA, USA, 1961.
23. Lehmann, E.L. *Theory of Point Estimation*; John Wiley & Sons: New York, NY, USA, 1983.
24. Nazin, A.V. On minimax bound for parameter estimation in ball (bias accounting). In *New Trends in Probability and Statistics*; Walter de Gruyter GmbH: Berlin, Germany, 1991; pp. 612–616.
25. Yang, Y.; Barron, A.R. Information-theoretic determination of minimax rates of convergence. *Ann. Stat.* **1999**, *27*, 1564–1599. [\[CrossRef\]](#)
26. Guntuboyina, A. *Minimax Lower Bounds*. Ph.D. Thesis, Statistics Department, Yale University, New Haven, CT, USA, 2011.
27. Guntuboyina, A. Lower bounds for the minimax risk using  $f$ -divergences, and applications. *IEEE Trans. Inform. Theory* **2011**, *57*, 2386–2399. [\[CrossRef\]](#)
28. Kim, A.K.H. Obtaining minimax lower bounds: A review. *J. Korean Stat. Soc.* **2020**, *49*, 673–701. [\[CrossRef\]](#)
29. Cover, T.M.; Thomas, J.A. *Elements of Information Theory*; John Wiley & Sons: Hoboken, NJ, USA, 2006.
30. Le Cam, L. *Asymptotic Methods in Statistical Decision Theory*; Springer: New York, NY, USA, 1986.
31. Yu, B.; Assouad, P.; Fano, R.M.; Le Cam, L. In *A Festschrift for Lucien Le Cam*; Pollard, D., Torgersen, E., Yang, G.L., Eds.; Springer: New York, NY, USA, 1997; pp. 423–435.
32. Cai, T.T.; Zhou, H.H. Optimal rates of convergence for sparse covariance matrix estimation. *Ann. Stat.* **2012**, *40*, 2389–2420. [\[CrossRef\]](#)

33. Helstrom, C.W. *Statistical Theory of Signal Detection*, 2nd ed.; Pergamon Press: Oxford, UK, 1975.
34. Whalen, D. *Detection of Signals in Noise*; Academic Press: New York, NY, USA, 1971.
35. Parag, P. Lecture 17: Minimax Bounds. In *Foundations of Machine Learning*; Indian Institute of Science: Bangalore, India, 2022. Available online: <https://ece.iisc.ac.in/~parimal/2022/ml/lecture-17.pdf> (accessed on 18 September 2024).
36. Rogers, J. Information Theoretic Minimax Lower Bounds. 2018. Available online: [https://jenniferbrennan.github.io/resources/Minimax\\_Lower\\_Bound\\_Notes.pdf](https://jenniferbrennan.github.io/resources/Minimax_Lower_Bound_Notes.pdf) (accessed on 18 September 2024).
37. Rigollet, P. Chapter 5: Minimax lower bounds. In *High Dimensional Statistics*; OpenCourseWare Lecture Notes; Massachusetts Institute of Technology: Cambridge, MA, USA, 2015. Available online: [https://ocw.mit.edu/courses/18-s997-high-dimensional-statistics-spring-2015/501374d1714bfd55ff6345189b9c2e26\\_MIT18\\_S997S15\\_Chapter5.pdf](https://ocw.mit.edu/courses/18-s997-high-dimensional-statistics-spring-2015/501374d1714bfd55ff6345189b9c2e26_MIT18_S997S15_Chapter5.pdf) (accessed on 18 September 2024).
38. Zhang, Z.; Shi, Z.; Gu, Y. Ziv–Zakai bound for DOAs estimation. *IEEE Trans. Signal Process.* **2022**, *71*, 136–149. [[CrossRef](#)]
39. Bell, K.L.; Steinberg, Y.; Ephraim, Y.; Van Trees, H.L. Extended Ziv–Zakai lower bound for vector parameter estimation. *IEEE Trans. Inform. Theory* **1997**, *43*, 624–637. [[CrossRef](#)]
40. Zhang, Z.; Shi, Z.; Zhou, C.; Yan, C.; Gu, Y. Ziv–Zakai bound for compressive time delay estimation. *IEEE Trans. Signal Process.* **2022**, *70*, 4006–4019. [[CrossRef](#)]
41. Sun, J.; Ma, S.; Xu, G.; Li, S. Trade-off between positioning and communication for millimeter wave systems with Ziv–Zakai bound. *IEEE Trans. Commun.* **2023**, *71*, 3752–3762. [[CrossRef](#)]
42. Mishra, K.V.; Eldar, Y.C. Performance of time delay estimation in a cognitive radar. In Proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), New Orleans, LA, USA, 5–9 March 2017; pp. 3141–3145.
43. Bell, K.L.; Ephraim, Y.; Van Trees, H.L. Explicit Ziv–Zakai lower bound for bearing estimation. *IEEE Trans. Signal Process.* **1996**, *44*, 2810–2824. [[CrossRef](#)]
44. Zhang, Z.; Shi, Z.; Gu, Y.; Greco, M.S.; Gini, F. Ziv–Zakai bound for compressive time delay estimation from zero-mean Gaussian signal. *Signal Process. Lett.* **2023**, *30*, 1112–1116. [[CrossRef](#)]
45. Chiriach, V.M.; He, Q.; Haimovich, A.M.; Blum, R.S. Ziv–Zakai bound for joint parameter estimation in MIMO radar systems. *IEEE Trans. Signal Process.* **2015**, *63*, 4956–4968. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.