Maximin Performance of Binary-Input Channels with Uncertain Noise Distributions

Andrew L. McKellips, *Student Member, IEEE*, and Sergio Verdú, *Fellow, IEEE*

***Abstract*—We consider uncertainty classes of noise distributions defined by a bound on the divergence with respect to a nominal noise distribution. The noise that maximizes the minimum error probability for binary-input channels is found. The effect of the reduction in uncertainty brought about by knowledge of the signal-to-noise ratio is also studied. The particular class of Gaussian nominal distributions provides an analysis tool for near-Gaussian channels. Asymptotic behavior of the least favorable noise distribution and resulting error probability are studied in a variety of scenarios, namely: asymptotically small divergence with and without power constraint; asymptotically large divergence with and without power constraint; and asymptotically large signal-to-noise ratio.**

***Index Terms*—Detection, Gaussian error probability, hypothesis testing, Kullback–Leibler divergence, least favorable noise.**

I. Introduction

ADDITIVE noise is most often represented by a fixed random variable, typically Gaussian, modeling the joint effect of such signal distortions as ambient channel noise, crosstalk, intersymbol interference and/or fading. In many instances, however, it may not be feasible or desirable to fix the exact noise distribution; instead, there might be a class of noise distributions that deserves analysis.

There is a rich literature in the area of worst case constrained noise and interference analysis, covering a wide range of transmission and interference strategies. For instance, worst case transition probabilities for a finite-input-alphabet, finite-output-alphabet channel are considered in [1], while assumed knowledge of the noise moments is used to generate a maximum-entropy distribution in [2]. Most often, a peak or average power constraint is imposed on the noise or interference and a worst case distribution is sought, either for guaranteed performance quantification [3]–[13] or in the context of a zero-sum game formulation between communicator and jammer [14], [15]. Worst case performance for power-constrained interference in a direct-sequence spread spectrum (DS/SSc) system with Gaussian background noise and linear matched-filter detector is considered in [3], with an extension to nonlinear detection considered in [4] for large spreading gain. Worst case power-constrained partial-band noise and multitone jamming for a variety of spread-spectrum transmission strategies are considered in [5]–[10]. Worst case amplitude- and power-constrained interference for an additive Gaussian channel with intersymbol interference is considered in [11]. Worst case power-constrained noise in "very noisy" channels is considered in [12] as an extension to [1]. A full solution for this problem is developed in [13] for maximum-likelihood detection, where the least favorable noise distributions for binary-input additive-noise channels with fixed signal-to-noise ratio (SNR) are found for several performance measures. An extension in [15] considers the zero-sum game between communicator and jammer when the communicator transmits an antipodal signal with pseudorandom amplitude pattern known only to the receiver.

In addition to, or in lieu of, a power constraint, a constraint on the proximity of the noise to a prescribed nominal distribution is interesting as it represents prior knowledge of the approximate behavior of the channel. Such knowledge is available, for instance, in a channel subject to dynamic perturbations, a system analysis which is computationally intensive, or a hypothesis test regarding the distribution of channel noise which involves decision classes determined by proximity bounds. Consider, for example, a linear multiuser detector which mitigates multiple-access interference in such a manner that the overall channel distortion, made up of multiple-access interference and background noise, resembles the noise that would prevail in the absence of interfering users but whose exact error probability is hard to determine. Such multiuser detectors have received much attention; for an overview see [16]. In [17], a bound is obtained for the (Kullback–Leibler) divergence between the distribution of multiple-access interference plus Gaussian channel noise and a nominal Gaussian distribution for the minimum-mean-square-error (MMSE) linear multiuser detector. It becomes of interest, then, to study the error probability characteristics of the noise class defined by a divergence (proximity) bound.

Worst case noise under constraints of power and divergence was studied in [18] for a zero-threshold detector. In this paper, we consider instead maximum-likelihood (ML) detection in order to quantify optimum detector performance in the presence of worst case noise, providing a best possible guaranteed performance level from the receiver's point of view. The ML detector problem is significantly more involved than the corresponding zero-threshold detection problem as a result of the dependence of the detector strategy on the noise distribution. Whereas the treatment in [18] involved a

Manuscript received November 15, 1996; revised November 23, 1997. This work was supported in part by the U.S. Army Research Office under Grant DAAH04-96-1-0379. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Ulm, Germany, July 1997.

The authors are with the Department of Electrical Engineering, School of Engineering and Applied Science, Princeton University, Princeton, NJ 08544 USA.

Publisher Item Identifier S 0018-9448(98)02358-X.

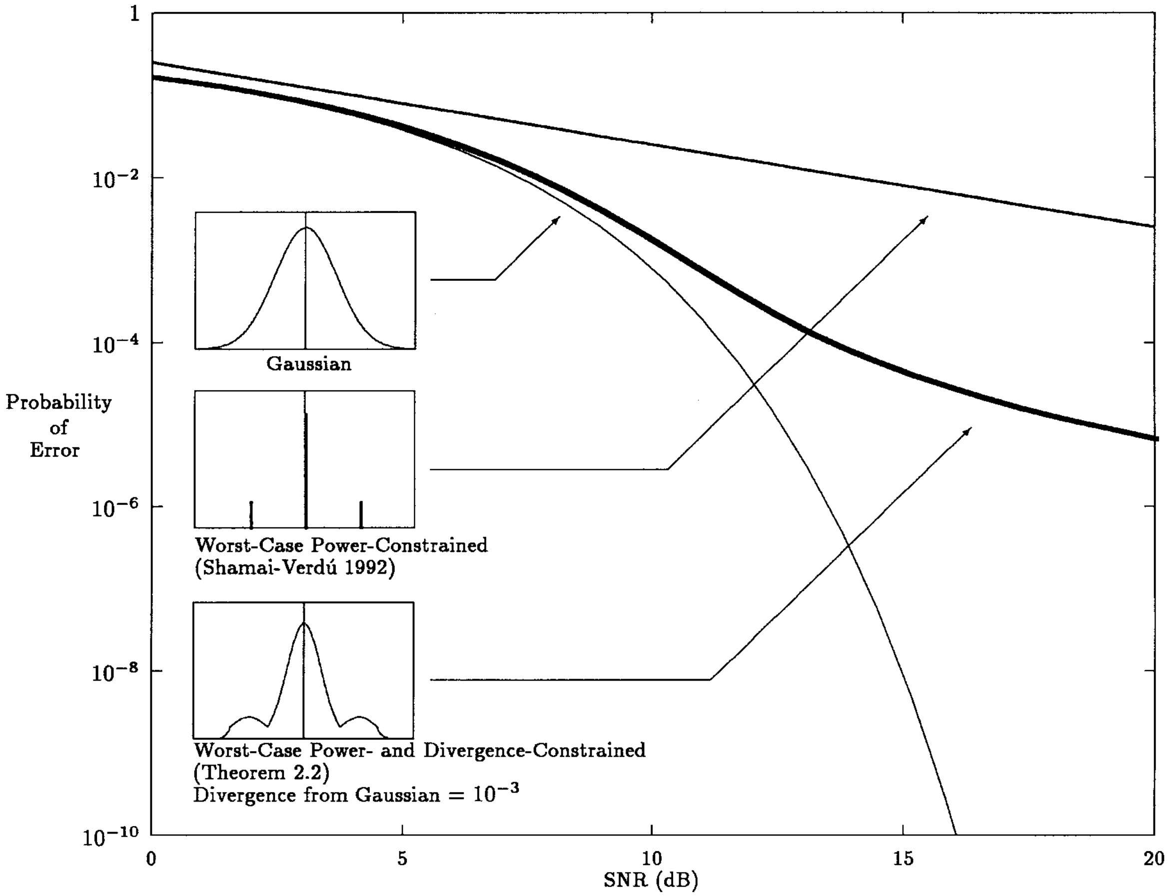

Fig. 1. Probability of maximum-likelihood detection error under various constraints.

Lagrange-multiplier optimization approach over fixed zero-threshold detection regions, such regions are unknown *a priori* in the ML detection formulation. Although zero-threshold detection is optimal for the least favorable distributions obtained in this paper, the least favorable noise distributions obtained are different from those of [18], implying that the maximin problem of interest does not have a saddle point.

Fig. 1 depicts the worst case error probability for a binary-input channel over a range of *SNR* values, with each curve representing a distinct set of constraints on the noise. The curve depicting standard Gaussian error probabilities corresponds with the constraining of the noise probability density function (pdf) to zero divergence with respect to a nominal Gaussian. The worst case power-constrained curve was obtained in [13] and corresponds to a total relaxation of any proximity constraint, including divergence. The worst case power- and divergence-constrained curve depicts worst case ML detection error probability subject to a finite, nonzero divergence-from-Gaussian constraint and is based on the results of this paper. As the divergence tolerance is decreased, this curve approaches the Gaussian performance curve, which represents a lower limit; as such tolerance is increased, the worst case power-constrained curve represents an upper limit, a result of Theorem 1.

In Section II, we develop an expression for the least favorable divergence-constrained noise, both with and without a power constraint. In Section III we focus on the particular class of Gaussian nominals. In Section IV, we study the asymptotic behavior of the solution in a variety of scenarios. Section V presents a summary of the results.

II. Worst Case Noise for Maximum-Likelihood Detection

A. Power and Divergence Constraints

Consider the decision hypothesis test for the standard binary-input additive-noise channel:

$$H_0\colon Y = -1 + \mathcal{N}$$
$$H_1\colon Y = +1 + \mathcal{N} \tag{1}$$

where $\mathcal{N}$ is a random variable representing additive channel noise. Fix a random variable N with symmetric probability density function (pdf) f_N exhibiting second moment

$$\sigma^2 = E[N^2] = \int_{-\infty}^{\infty} x^2 f_N(x)\, dx$$

and consider a class of pdf's constrained in power and divergence with respect to f_N. Within this class, we wish to find a pdf $f_{\mathcal{N}}$ which maximizes the ML probability of detection error criterion

$$\begin{aligned} P_{\mathrm{ML}}(f_{\mathcal{N}}) &= \tfrac{1}{2} \int_{-\infty}^{\infty} \min\{f_{\mathcal{N}}(x-1), f_{\mathcal{N}}(x+1)\}\, dx \\ &= \tfrac{1}{2} \int_{-\infty}^{\infty} \min\{f_{\mathcal{N}}(x), f_{\mathcal{N}}(x+2)\}\, dx \end{aligned} \tag{2}$$

corresponding to (1), where the random variable $\hat{N}$ with pdf $f_{\hat{N}}$ represents worst case power- and divergence-constrained noise.

The divergence of any pdf $f_{\mathcal{N}}$ with respect to a nominal f_N is defined by

$$D(f_{\mathcal{N}} \| f_N) = \int_{-\infty}^{\infty} f_{\mathcal{N}}(x) \log \frac{f_{\mathcal{N}}(x)}{f_N(x)}\, dx$$

whenever $f_{\mathcal{N}}$ exhibits absolute continuity with respect to f_N, and is otherwise set by convention to infinity (see [19], for instance); we take $\log(\cdot)$ to represent the natural (base e) logarithm. We assume, unless otherwise noted, that the nominal f_N is symmetric continuous with support

$$\{x: f(x) > 0\} = (-a, a).$$

We note that the class of worst case candidate pdf's is restricted in support to the same interval $(-a, a)$ through absolute continuity for any finite-divergence tolerance. Since any pdf with mass restricted to the interval $(-1, 1)$ exhibits zero ML probability of error according to (2), the problem is trivial unless $a > 1$. Relaxation of the continuity restriction on f_N is discussed in Section II-C.

Explicitly, the described optimization problem is given by

$$f_{\hat{N}} = \arg \max_{f_{\mathcal{N}}} P_{\mathrm{ML}}(f_{\mathcal{N}}) \tag{3}$$

subject to the constraints

$$\int_{-a}^{a} f_{\mathcal{N}}(x)\, dx = 1 \tag{4}$$

$$\int_{-a}^{a} x^2 f_{\mathcal{N}}(x)\, dx \le \sigma^2 \tag{5}$$

$$D(f_{\mathcal{N}} \| f_N) \le \delta \tag{6}$$

where, for a general set Y and real-valued function $g(\cdot)$ with domain containing Y, we take $\arg\max_Y g(Y)$ to be any y^* in Y for which $g(y^*) \ge g(y)$ for all y in Y. Note that the divergence tolerance δ can take values in $(0, \infty)$ since divergence is always nonnegative and the case $\delta = 0$ is trivially realized by $f_{\hat{N}} \equiv f_N$.

We take all integrals to be with respect to Lebesgue measure. We also adopt the convention that any pdf f actually denotes the equivalence class of all pdf's agreeing with f up to a set of measure zero, since such discrepancies affect neither the objective function nor the constraints in (3). The ML probability of detection error $P_{\mathrm{ML}}(f_{\hat{N}})$ achieved by the worst case noise $\hat{N}$ is often of as much interest as the form of $f_{\hat{N}}$ itself, and will be denoted by $P_{\hat{N}}$. The augmented notation $f_{\hat{N},N,\delta}$ and $P_{\hat{N},N,\delta}$ will be used to represent $f_{\hat{N}}$ and $P_{\hat{N}}$, respectively, whenever dependence on the nominal f_N and divergence tolerance δ deserve particular attention.

An important observation regarding the optimization problem (3) concerns convexity of the feasible set. Given any nominal f_N and any two distinct feasible pdf's f_1 and f_2, the candidate pdf f_3 defined by

$$f_3(x) = \alpha\, f_1(x) + (1-\alpha) f_2(x), \qquad 0 < \alpha < 1$$

clearly satisfies (4) and (5). Furthermore, the divergence functional $D(\cdot \| f_N)$ is strictly convex (see [19, p. 30] for instance). The objective function P_{ML} is concave as demonstrated for any two pdf's f_1, f_2 and any $0 < \alpha < 1$ by the following relationship:

$$\begin{aligned}
&P_{\mathrm{ML}}(\alpha f_1 + (1-\alpha) f_2) \\
&\quad = \frac{1}{2} \int_{-a}^{a} \min\{\alpha f_1(x) + (1-\alpha) f_2(x), \\
&\qquad\qquad \alpha f_1(x+2) + (1-\alpha) f_2(x+2)\}\, dx \\
&\quad \ge \frac{\alpha}{2} \int_{-a}^{a} \min\left\{ f_1(x), f_1(x+2)\right\} dx \\
&\qquad\qquad + \frac{1-\alpha}{2} \int_{-a}^{a} \min\{f_2(x), f_2(x+2)\left.\right\} dx \\
&\quad = \alpha P_{\mathrm{ML}}(f_1) + (1-\alpha) P_{\mathrm{ML}}(f_2).
\end{aligned}$$

Note that, as opposed to the divergence measure, concavity is not necessarily strict in the objective function.

The first observation about the solution set for (3) concerns uniqueness of the solution. The proof is deferred to the Appendix, along with the proofs for Lemmas 2–5, 7–9, 15, 16, and Theorem 1. For either of the inequality constraints (5) and (6), the constraint will be considered active whenever optimality of (3) is achieved only by a solution or solutions for which the constraint is met with equality.

Lemma 1: Given a continuous symmetric nominal f_N with support $(-a, a)$, the solution $f_{\hat{N}}$ to (3) is unique whenever the divergence constraint is active.

The second observation concerns symmetry of the solution.

Lemma 2: Given a continuous symmetric nominal pdf f_N with support $(-a, a)$, the solution $f_{\hat{N}}$ to (3) is symmetric whenever the divergence constraint is active.

We now turn our attention to characterizing the conditions under which the divergence constraint is active. The approach we take makes use of the asymptotic behavior of the solution $f_{\hat{N},N,\delta}$ as the divergence tolerance δ grows unbounded, which is of independent interest. The worst case power-constrained solution corresponding to (3), without an imposed divergence constraint (i.e., without reference to a nominal noise distribution), was shown in [13] to be realized by a probability mass function $\mathcal{P}_{\sigma^2}^{SV}$ taking mass on the set

$$\{-M_{\sigma^2}, -M_{\sigma^2}+1, \cdots, M_{\sigma^2}\}$$

where M_{σ^2} is a positive integer given by

$$M_{\sigma^2} = \lceil \sqrt{3\sigma^2 + 1} - 1 \rceil$$

with σ^2 representing the noise power. It was additionally shown that $\mathcal{P}_{\sigma^2}^{SV}$ is a mixture of two equiprobable distributions, one taking the value p_1 on the span-2 lattice $\{-M_{\sigma^2}, -M_{\sigma^2}+2, \cdots, M_{\sigma^2}\}$ and the other the value p_2 on the span-2 lattice $\{-M_{\sigma^2}+1, -M_{\sigma^2}+3, \cdots, M_{\sigma^2}-1\}$ where $p_1 \ge 0, p_2 \ge 0$, and $(M_{\sigma^2}+1)p_1 + M_{\sigma^2} p_2 = 1$.

We might intuitively expect the solution $f_{\hat{N},N,\delta}$ to (3) to approach $\mathcal{P}_{\sigma^2}^{SV}$ as we allow the divergence tolerance to tend to infinity regardless of the prescribed nominal f_N. However, we immediately discount this tendency for nominals without mass at any of the points $-M_{\sigma^2}, -M_{\sigma^2}+1, \cdots, M_{\sigma^2}$ since finite divergence requires absolute continuity of the worst case pdf $f_{\hat{N}}$ with respect to f_N. We nonetheless expect the tendency to hold for nominals with mass at these points, and prove the following result in the Appendix.

Theorem 1: Given a nominal f_N with second moment σ^2 and support $(-a, a)$ where

$$a > \lceil \sqrt{3\sigma^2+1} - 1 \rceil$$

the solution to (3) satisfies

$$\lim_{\delta \to \infty} F_{\hat{N},N,\delta} = F_{\sigma^2}^{SV}$$

where $F_{\hat{N},N,\delta}$ and $F_{\sigma^2}^{SV}$ represent the cumulative distribution functions of $f_{\hat{N},N,\delta}$ and $\mathcal{P}_{\sigma^2}^{SV}$, respectively, and where the limit corresponds to convergence in distribution, defined by pointwise convergence everywhere except perhaps at points of discontinuity.

Theorem 1 demonstrates convergence in distribution of worst case power- and divergence-constrained noise to worst case power-constrained noise with growing divergence tolerance for pdf's with sufficient single-interval support. It is worth noting that the result can be easily generalized to encompass all nominals f_N with second moment σ^2 and support that includes intervals of positive measure around each of the points $-M_{\sigma^2}, -M_{\sigma^2}+1, \cdots, M_{\sigma^2}$ where $M_{\sigma^2} = \lceil \sqrt{3\sigma^2+1} - 1 \rceil$; this is the most general class of nominal pdf for which the result holds.

The class of nominal pdf's treated in Theorem 1 encompasses most typical channel models, since channel noise is most often assumed to have a symmetric pdf with infinite support, for instance a Gaussian distribution. Moreover, if we consider nominals with finite support and satisfying the typical additional assumption of unimodality, where we take any symmetric pdf f to be unimodal if $f(x)$ is monotone decreasing for $x > 0$ in accordance with [20, p. 158], then asymptotic behavior of those nominals with support $(-a, a)$ where a falls in the set

$$S = \bigcup_{i=1}^{\infty} [i, \sqrt{(i+1)^2 - 1})$$

is also governed by Theorem 1; justification stems from the observation that the second moment of such a distribution is bounded above by that of the uniform distribution over the interval $[-a, a]$, given by $a^2/3$, while

$$\lceil \sqrt{3(a^2/3)+1} - 1 \rceil = \lceil \sqrt{a^2+1} - 1 \rceil < a$$

for all a in S. Nominals which do not fall into the class governed by Theorem 1 can be subjected to a negligible perturbation by transporting small quantities of mass to form new intervals and/or tails in order to construct a new nominal which does. An exception to this idea arises when intervals of zero mass in the pdf represent intentional restrictions on potential noise values; for example, the choice of a nominal f_N with mass restricted to a finite interval $(-a, a)$ effectively constrains the candidate noise in amplitude. It is of interest, then, to analyze the counterpart scenario to Theorem 1, when the nominal has support $(-a, a)$ with $a \leq M_{\sigma^2}$.

In the case $a \leq M_{\sigma^2}$, it is clear that the finite support restriction on $f_{\hat{N}}$ derived from the absolute continuity requirement of the divergence measure becomes more restrictive than the power constraint in (3) for large values of δ. Hence, in order to characterize the asymptotic behavior of the solution, we make use of the upper bound on performance provided by worst case amplitude-constrained rather than worst case power-constrained noise. Such a problem was studied in detail in [21], where it was shown that worst case amplitude-constrained noise for an amplitude bound a is achieved by a class $\mathcal{M}(a)$ of "picket-fence" distributions consisting of 2-periodic pdf's (i.e., periodic with period 2) with support restricted to $[-a, a]$ and which satisfy the requirement that every interval of positive mass exhibit $\lfloor a \rfloor + 1$ periods. Such distributions collectively exhibit ML probability of detection error $\lfloor a \rfloor / (2(\lfloor a \rfloor + 1))$, providing an upper bound for the problem at hand. This result leads to the following characterization.

Lemma 3: Given a continuous symmetric nominal f_N with second moment σ^2 and support $(-a, a)$ where

$$a \leq \lceil \sqrt{3\sigma^2+1} - 1 \rceil$$

there exists $\delta_{N,\sigma^2} < \infty$ such that the set of solutions to (3) falls within the class $\mathcal{M}(a)$ for all $\delta \geq \delta_{N,\sigma^2}$.

Armed with a description of the limiting behavior of $f_{\hat{N}M,\delta}$ as $\delta \to \infty$, we now characterize activity of the divergence constraint by making use of the convexity of the feasible set.

Lemma 4: Given a nominal f_N with second moment σ^2 and support $(-a, a)$, activity of the divergence constraint (6) in the optimization problem (3) satisfies the following. If $a > \lceil \sqrt{3\sigma^2+1} - 1 \rceil$, the divergence constraint is active for all δ. If $a \leq \lceil \sqrt{3\sigma^2+1} - 1 \rceil$, the divergence constraint is active for all $\delta \leq \delta_{N,\sigma^2}$ where

$$\delta_{N,\sigma^2} = \min_{f_{\mathcal{N}} \in \mathcal{M}(a):\ \int_{-a}^{a} x^2 f_{\mathcal{N}}(x)\, dx \leq \sigma^2} D(f_{\mathcal{N}} \| f_N).$$

The standing assumption throughout that the nominal pdf f_N has support on the whole interval $(-a, a)$ holds for typical channel models and is conducive to the development of general results. Indeed, asymptotic behavior of worst case noise and a characterization of activity of the divergence constraint have been established for such nominals earlier in this section. In the unusual event that the support of the nominal cosists of disjoint intervals (requiring, for instance, that the nominal pdf not exhibit unimodality), and for which the transportation of negligible quantities of mass in order to form a single interval of support as discussed above is not warranted, the following observation can aid in simplifying the analysis of worst case noise.

Lemma 5: Any solution to the optimization problem obtained by replacing the divergence constraint (6) with the corresponding equality constraint

$$D(f_{\mathcal{N}} \| f_N) = \delta$$

is also a solution to (3).

Lemma 5 permits a simplified analysis using Lagrange multiplier techniques even in cases where the divergence constraint is not necessarily active in the optimization problem (3).

The next observation concerns continuity of the solution.

Lemma 6: Given a continuous symmetric nominal f_N with support $(-a, a)$, the solution $f_{\hat{N}}$ to (3) is continuous whenever the divergence constraint is active.

Proof: This result is shown to hold in the Appendix for nominals f_N and values of δ for which there exists $\delta_2 > \delta$ such that

$$P_{\mathrm{ML}}(f_{\hat{N}N,\delta_2}) > P_{\mathrm{ML}}(f_{\hat{N}N,\delta}).$$

According to Lemma 4, such is the case for all values of δ whenever $a > M_{\sigma^2}$, and for all $\delta < \delta_{N,\sigma^2}$ whenever $a \le M_{\sigma^2}$. Since the divergence constraint is inactive for all $\delta > \delta_{N,\sigma^2}$ whenever $a \le M_{\sigma^2}$, the result is therefore shown to hold under the given assumptions except perhaps in the particular case $a \le M_{\sigma^2}$, $\delta = \delta_{N,\sigma^2}$ for which the result is shown to hold in Section IV-D. □

This result will prove useful in characterizing the form of $f_{\hat{N}}$ once a piecewise description has been established in a later result. We can further conclude from the proof of Lemma 6 that there exists a continuous pdf among the solutions to (3) even when the divergence constraint is not active.

Equipped with Lemmas 1, 2, 4, and 6 characterizing worst case noise for continuous symmetric nominals with a single interval of support, we now concentrate on developing a more detailed description of worst case noise. The following result for symmetric unimodal nominals provides some insight into the form of worst case noise, in addition to providing the basis for a subsequent result characterizing activity of the power constraint.

Lemma 7: Given a symmetric unimodal nominal f_N, the zero-threshold detector is an ML detector for the worst case noise pdf $f_{\hat{N}}$.

Note that unimodality of $f_{\hat{N},N,\delta}$ is not a necessary implication of Lemma 7. Nor is it implied that the zero-threshold detector falls in the ML class for all candidate noise distributions for (3), but does so for at least the nominal and worst case distributions.

We now turn our attention to characterizing the worst case pdf $f_{\hat{N}}$, making use of the results developed thus far. A Lagrange-multiplier analysis similar to that carried through in [18] would be feasible were it not for the nonsmoothness of the $\min\{\cdot,\cdot\}$ functional in the objective function (2).

The form of (2) suggests the following partitioning of the support $(-a, a)$ of f_N given any candidate pdf $f_{\mathcal{N}}$:

$$\begin{aligned}
\mathcal{A} &= \{x \in (-a,a)\colon f_{\mathcal{N}}(x-2) < f_{\mathcal{N}}(x) < f_{\mathcal{N}}(x+2)\}\\
&\quad \cup\ \{x \in (-a,a)\colon f_{\mathcal{N}}(x-2) > f_{\mathcal{N}}(x) > f_{\mathcal{N}}(x+2)\}\\
\mathcal{B} &= \{x \in (-a,a)\colon f_{\mathcal{N}}(x-2) < f_{\mathcal{N}}(x) \text{ and }\\
&\quad f_{\mathcal{N}}(x) > f_{\mathcal{N}}(x+2)\}\\
\overline{\mathcal{B}} &= \{x \in (-a,a)\colon f_{\mathcal{N}}(x-2) > f_{\mathcal{N}}(x) \text{ and }\\
&\quad f_{\mathcal{N}}(x) < f_{\mathcal{N}}(x+2)\}\\
\mathcal{C} &= \{x \in (-a,a)\colon f_{\mathcal{N}}(x-2) = f_{\mathcal{N}}(x)\}\\
&\quad \cup\ \{x \in (-a,a)\colon f_{\mathcal{N}}(x) = f_{\mathcal{N}}(x+2)\}.
\end{aligned}$$

Now further categorize the set $\mathcal{C}$ by defining for $n = 2, 3, \cdots$

$$\begin{aligned}
\mathcal{C}_n &= \{x \in \mathcal{C}\colon f_{\mathcal{N}}(x-2n) < f_{\mathcal{N}}(x-2n+2) \text{ and }\\
&\quad f_{\mathcal{N}}(x-2n+2) = f_{\mathcal{N}}(x-2n+4) = \cdots = f_{\mathcal{N}}(x)\\
&\quad \text{and } f_{\mathcal{N}}(x) > f_{\mathcal{N}}(x+2)\}\\
\bar{\mathcal{C}}_n &= \{x \in \mathcal{C}\colon f_{\mathcal{N}}(x-2n) > f_{\mathcal{N}}(x-2n+2) \text{ and }\\
&\quad f_{\mathcal{N}}(x-2n+2) = f_{\mathcal{N}}(x-2n+4) = \cdots = f_{\mathcal{N}}(x)\\
&\quad \text{and } f_{\mathcal{N}}(x) < f_{\mathcal{N}}(x+2)\}\\
\mathcal{C}_n^L &= \{x \in \mathcal{C}\colon f_{\mathcal{N}}(x-2n) < f_{\mathcal{N}}(x-2n+2)\\
&\quad = f_{\mathcal{N}}(x-2n+4) = \cdots = f_{\mathcal{N}}(x) < f_{\mathcal{N}}(x+2)\}\\
\mathcal{C}_n^R &= \{x \in \mathcal{C}\colon f_{\mathcal{N}}(x-2n) > f_{\mathcal{N}}(x-2n+2)\\
&\quad = f_{\mathcal{N}}(x-2n+4) = \cdots = f_{\mathcal{N}}(x) > f_{\mathcal{N}}(x+2)\}
\end{aligned}$$

with the limiting conventions

$$\begin{aligned}
\mathcal{C}_\infty^L &= \{x \in \mathcal{C}\colon f_{\mathcal{N}}(x+2) > f_{\mathcal{N}}(x) = 0\\
&\quad = f_{\mathcal{N}}(x-2) = f_{\mathcal{N}}(x-4) = \cdots\},\\
\mathcal{C}_\infty^R &= \{x \in \mathcal{C}\colon f_{\mathcal{N}}(x-2) > f_{\mathcal{N}}(x) = 0\\
&\quad = f_{\mathcal{N}}(x+2) = f_{\mathcal{N}}(x+4) = \cdots\},\\
\mathcal{C}_\infty &= \{x \in \mathcal{C}, x \in [0,2)\colon \cdots = f_{\mathcal{N}}(x-2)\\
&\quad = f_{\mathcal{N}}(x) = 0 = f_{\mathcal{N}}(x+2) = \cdots\}.
\end{aligned}$$

For example, any Gaussian nominal pdf with mean μ exhibits the support decomposition

$$\begin{aligned}
\mathcal{A} &= (-\infty, \mu-1) \cup [\mu+1, \infty)\\
\mathcal{B} &= [\mu-1, \mu+1)
\end{aligned}$$

with all other sets being empty. The support decomposition for the pdf depicted in Fig. 2 is given by

$$\mathcal{C}_4 = [5/2, 7/2), \mathcal{C}_\infty = [0, 1/2) \cup [3/2, 2)$$

with all remaining sets empty.

Denote by Λ the collection

$$\{\mathcal{A}, \mathcal{B}, \overline{\mathcal{B}}, \mathcal{C}_2, \overline{\mathcal{C}}_2, \mathcal{C}_2^L, \mathcal{C}_2^R, \mathcal{C}_3, \overline{\mathcal{C}}_3, \mathcal{C}_3^L, \mathcal{C}_3^R, \cdots\}$$

and by $\mathcal{F}_\Lambda$ the set of pdf's which exhibit the decomposition Λ. Note that, while the collection $\{\mathcal{A}, \mathcal{B}, \overline{\mathcal{B}}, \mathcal{C}\}$ forms a partitioning of the support $(-a, a)$, for nonempty $\mathcal{C}$ the collection Λ does not. However, any candidate $f_{\mathcal{N}}$ *is* completely determined by its specification over the sets of Λ as a result of the definitions of $\mathcal{C}_n, \overline{\mathcal{C}}_n, \mathcal{C}_n^L$, and $\mathcal{C}_n^R$. The purpose of this construction is to restrict attention to the subset of the support $(-a, a)$ over which the objective function is smooth by grouping the singular points x for which $f_{\mathcal{N}}(x) = f_{\mathcal{N}}(x-2)$ or $f_{\mathcal{N}}(x) = f_{\mathcal{N}}(x+2)$ (i.e., those x in $\mathcal{C}$) in order to submit to a Lagrange-multiplier analysis. There is no loss of information in this grouping since the grouped points are fully reconstructable by the definition determining their candidacy for grouping; for instance, knowledge of $f_{\mathcal{N}}(x)$ over the set $\mathcal{C}_n$ completely determines $f_{\mathcal{N}}(x)$ for all x for which $x+2i \in \mathcal{C}_n, i = 1, 2, \cdots, n-1$. This permits full expression of the objective function (2) and the constraints (4)–(6) over the sets of Λ. Using this idea, the original optimization problem

(3) can be written in a decoupled form. In this framework, the worst case pdf $f_{\hat{N}}$ is the (not necessarily unique) pdf for which

$$\begin{aligned} P_{\rm ML}(f_{\hat{\mathcal{N}}}) &= \max_{f_{\mathcal{N}}} \; P_{\rm ML}(f_{\mathcal{N}}) \\ &= \max_{\Lambda} \; \max_{f_{\mathcal{N}} \in \mathcal{F}_{\Lambda}} \; P_{\rm ML}(f_{\mathcal{N}}) \end{aligned} \tag{7}$$

where

$$\begin{aligned} P_{\rm ML}(f_{\mathcal{N}}) = \tfrac{1}{2} \int^{\mathcal{A}} f_{\mathcal{N}}(x)\,dx + \int_{\overline{\mathcal{B}}} f_{\mathcal{N}}(x)\,dx \\ + \frac{1}{2} \sum_{n=2}^{\infty} \Bigg[(n-1) \int_{\mathcal{C}_n} f_{\mathcal{N}}(x)\,dx \\ + n \int_{\mathcal{C}_n^L \cup \mathcal{C}_n^R} f_{\mathcal{N}}(x)\,dx \\ + (n+1) \int_{\overline{\mathcal{C}}_n} f_{\mathcal{N}}(x)\,dx \Bigg] \end{aligned}$$

subject to the constraints

$$\int_{\mathcal{A} \cup \mathcal{B} \cup \overline{\mathcal{B}}} f_{\mathcal{N}}(x)\,dx + \sum_{n=2}^{\infty} n \int_{\mathcal{C}_n \cup \overline{\mathcal{C}}_n \cup \mathcal{C}_n^L \cup \mathcal{C}_n^R} f_{\mathcal{N}}(x)\,dx = 1 \tag{8}$$

$$\begin{aligned} \int_{\mathcal{A} \cup \mathcal{B} \cup \overline{\mathcal{B}}} x^2 f_{\mathcal{N}}(x)\,dx \\ + \sum_{n=2}^{\infty} \int_{\mathcal{C}_n \cup \overline{\mathcal{C}}_n \cup \mathcal{C}_n^L \cup \mathcal{C}_n^R} \sum_{i=0}^{n-1} (x-2i)^2 f_{\mathcal{N}}(x)\,dx \le \sigma^2 \end{aligned} \tag{9}$$

$$\begin{aligned} \int_{\mathcal{A} \cup \mathcal{B} \cup \overline{\mathcal{B}}} f_{\mathcal{N}}(x) \log \frac{f_{\mathcal{N}}(x)}{f_N(x)}\,dx \\ + \sum_{n=2}^{\infty} \int_{\mathcal{C}_n \cup \overline{\mathcal{C}}_n \cup \mathcal{C}_n^L \cup \mathcal{C}_n^R} \sum_{i=0}^{n-1} f_{\mathcal{N}}(x) \; \log \; \frac{f_{\mathcal{N}}(x)}{f_N(x-2i)}\,dx \\ \le \delta. \end{aligned} \tag{10}$$

The worst case decomposition represented by the outer optimization of (7) corresponds to an optimal choice of a set of labelled intervals within the support. The Lagrangian associated with the inner optimization of (7) is given by

$$\begin{aligned} L(x; f_N, &\Lambda, \lambda_1, \lambda_2, \lambda_3) \\ = &\tfrac{1}{2}\, 1_{\mathcal{A}}(x) + 1_{\overline{\mathcal{B}}}(x) \\ &+ \frac{1}{2} \sum_{n=2}^{\infty} ((n-1) 1_{\mathcal{C}_n}(x) + n 1_{\mathcal{C}_n^L \cup \mathcal{C}_n^R}(x) \\ &+ (n+1) 1_{\overline{\mathcal{C}}_n}(x)) \\ &+ \lambda_1 \left(1_{\mathcal{A} \cup \mathcal{B} \cup \overline{\mathcal{B}}}(x) + \sum_{n=2}^{\infty} n 1_{\mathcal{C}_n \cup \mathcal{C}_n^L \cup \mathcal{C}_n^R \cup \overline{\mathcal{C}}_n}(x) \right) \\ &+ \lambda_2 \Bigg(x^2 1_{\mathcal{A} \cup \mathcal{B} \cup \overline{\mathcal{B}}}(x) + \sum_{n=2}^{\infty} 1_{\mathcal{C}_n \cup \mathcal{C}_n^L \cup \mathcal{C}_n^R \cup \overline{\mathcal{C}}_n}(x) \\ &\cdot \sum_{i=0}^{n-1} (x-2i)^2 \Bigg) \\ &+ \lambda_3 \Bigg(1_{\mathcal{A} \cup \mathcal{B} \cup \overline{\mathcal{B}}}(x) + \left(1 + \log \frac{f_{\mathcal{N}}(x)}{f_N(x)} \right) \\ &+ \sum_{n=2}^{\infty} 1_{\mathcal{C}_n \cup \mathcal{C}_n^L \cup \mathcal{C}_n^R \cup \overline{\mathcal{C}}_n}(x) \\ &\cdot \sum_{i=0}^{n-1} \left(1 + \log \frac{f_{\mathcal{N}}(x)}{f_N(x-2i)} \right) \Bigg) \end{aligned}$$

where $1_Y(\cdot)$ represents the indicator function for the set Y. Recalling Lemma 5, which allows for the substitution of an equality constraint on divergence, the worst case pdf $f_{\hat{N}}$ reduces without loss of generality to the parametric form

$$f_{\hat{N}}(x) = \begin{cases} k_1 f_{\bar{N}}(x), & x \in \mathcal{A} \\ k_0 f_{\bar{N}}(x), & x \in \mathcal{B} \\ \overline{k}_0 f_{\bar{N}}(x), & x \in \overline{\mathcal{B}} \\ k_n \left(\prod_{i=0}^{n-1} f_{\bar{N}}(x-2i) \right)^{\frac{1}{n}}, & \\ \qquad x \in \mathcal{C}_n, n = 2, 3, \cdots & \\ \overline{k_n} \left(\prod_{i=0}^{n-1} f_{\bar{N}}(x-2i) \right)^{\frac{1}{n}}, & \\ \qquad x \in \overline{\mathcal{C}}_n, n = 2, 3, \cdots & \\ k_n^* \left(\prod_{i=0}^{n-1} f_{\bar{N}}(x-2i) \right)^{\frac{1}{n}}, & \\ \qquad x \in \mathcal{C}_n^L \cup \mathcal{C}_n^R, n = 2, 3, \cdots & \end{cases} \tag{11}$$

where

$$f_{\bar{N}}(x) = C f_N(x) e^{-cx^2} \tag{12}$$

and where $c, C, k_0, \overline{k}_0, k_1, k_2, \overline{k}_2, k_2^*, k_3, \overline{k}_3, k_3^*, \cdots$ are constants. Note that the constant C does not represent a degree of freedom as it is determined by the value of c and the constraint that $f_{\bar{N}}$ have unit integral. We will refer to $f_{\bar{N}}$ as the *variance-scaled nominal* corresponding to the nominal f_N.

Of note is the fact that any specification of any one of the multiplicative constants $k_0, k_1, k_2, k_2^*, \cdots$ fully determines the others through continuity whenever the divergence constraint is active according to Lemma 6.

In turn, the single remaining degree of freedom is fully determined by the constraint (8). Hence, the optimization problem (7) has been reduced to the determination of an optimal value of c and an optimal support decomposition Λ. Our goal is to reduce the choice of labeled intervals representing Λ to one parameter, yielding a parametric expression for the worst case pdf $f_{\hat{N}}$ with two degrees of freedom associated with the constraints (9) and (10). This desired reduction is the result of the following Lemma.

Lemma 8: Given a continuous symmetric unimodal variance-scaled nominal $f_{\bar{N}}$ with support $(-a, a)$ and satisfying

$$f_{\bar{N}}(n+\tau_1) f_{\bar{N}}(n-\tau_1) > f_{\bar{N}}(n+\tau_2) f_{\bar{N}}(n-\tau_2) \tag{13}$$

for all integers $n \in (-a, a)$ and all

$$0 \le \tau_1 < \tau_2 < \min\{1, a - |n|\}$$

the optimal form of support decomposition Λ in (7) is completely determined by a single parameter $\overline{x} \in (1, a)$ whenever

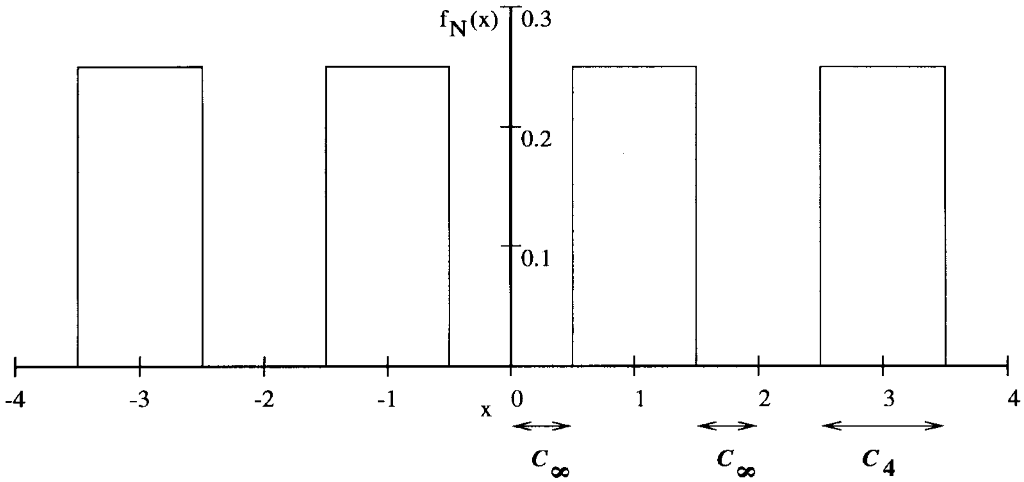

Fig. 2. Example of a pdf exhibiting support decomposition $\mathcal{C}_4 = [5/2, 7/2)$, $\mathcal{C}_\infty = [0, 1/2) \cup [3/2, 2)$.

the divergence constraint is active, and is given by

$$\Lambda = \begin{cases} \{\mathcal{A}, \mathcal{B}, \mathcal{C}_2\}, & \lfloor x \rfloor = 1 \\ \{\mathcal{A}, \mathcal{C}_N, \mathcal{C}_{N+1}\}, & \lfloor x \rfloor \geq 2 \end{cases} \tag{14}$$

where

$$\begin{aligned} \mathcal{A} &= (-a, -\overline{x}) \cup [\overline{x}, a) \\ \mathcal{B} &= [-2 + \overline{x}, 2 - \overline{x}), \lfloor x \rfloor = 1 \\ \mathcal{C}_{\lfloor x \rfloor} &= [\overline{x} - 2, 2\lfloor x \rfloor - \overline{x}), \lfloor x \rfloor \geq 2 \\ \mathcal{C}_{\lfloor x \rfloor + 1} &= [2\lfloor x \rfloor - \overline{x}, \overline{x}), \lfloor x \rfloor \geq 2. \end{aligned}$$

Note that the worst case pdf $f_{\hat{N}}$ is 2-periodic within the interval $[-\overline{x}, \overline{x})$.

An important observation is that if

$$f_N(n + \tau_1) f_N(n - \tau_1) > f_N(n + \tau_2) f_N(n - \tau_2) \tag{15}$$

for all integers $n \in (-a, a)$ and all

$$0 \leq \tau_1 < \tau_2 < \min\{1, a - |n|\}$$

then condition (15) holds for all $f_{\tilde{N}}$. Condition (15) amounts to a weak form of local log-concavity about the integers. Clearly, a sufficient condition for (15) is provided by the stronger requirement that

$$f_N(x + \tau_1) f_N(x - \tau_1) > f_N(x + \tau_2) f_N(x - \tau_2) \tag{16}$$

for all $x \in (-a, a)$ and all

$$0 \leq \tau_1 < \tau_2 < a - |x|$$

which is equivalent to the requirement that f_N be strictly log-concave over its support. This provides an easy check for many nominals, such as Gaussians.

Combining our previous analysis with Lemma 8 produces the following theorem characterizing worst case power- and divergence-constrained noise for ML detection.

Theorem 2: Given a continuous symmetric unimodal nominal f_N with second moment σ^2 and support $(-a, a)$, where either $a > \lceil \sqrt{3\sigma^2 + 1} - 1 \rceil$ or both $a \leq \lceil \sqrt{3\sigma^2 + 1} - 1$ and $\delta \leq \delta_{N,\sigma^2}$, and satisfying (15), the solution to (3) is given by

$$f_{\hat{N}}(x) = \begin{cases} k_0 f_{\tilde{N}}(x), \quad |x| > \overline{x} \\ k_{\lfloor \overline{x} \rfloor} \left(\displaystyle\prod_{i=0}^{\lfloor \overline{x} \rfloor - 1} f_{\tilde{N}}(x - 2i) \right)^{\frac{1}{\lfloor \overline{x} \rfloor}}, \\ \qquad x \in [\overline{x} - 2, 2\lfloor \overline{x} \rfloor - \overline{x}] \\ k_{\lfloor \overline{x} \rfloor + 1} \left(\displaystyle\prod_{i=0}^{\lfloor \overline{x} \rfloor} f_{\tilde{N}}(x - 2i) \right)^{\frac{1}{\lfloor \overline{x} \rfloor + 1}}, \\ \qquad x \in [2\lfloor \overline{x} \rfloor - \overline{x}, \overline{x}] \\ f_{\hat{N}}(x + 2i), \\ \qquad x + 2i \in [\overline{x} - 2, 2\lfloor \overline{x} \rfloor - \overline{x}], i = 1, 2, \cdots, \\ \qquad\qquad \lfloor \overline{x} \rfloor - 1 \\ f_{\hat{N}}(x + 2i), \\ \qquad x + 2i \in [2\lfloor \overline{x} \rfloor - \overline{x}, \overline{x}], i = 1, 2, \cdots, \lfloor \overline{x} \rfloor \end{cases} \tag{17}$$

where $\overline{x} \in (1, a)$ and

$$f_{\tilde{N}}(x) = C f_N(x) e^{-cx^2} \tag{18}$$

C being chosen to make $f_{\tilde{N}}$ a proper pdf.

Proof: The result follows directly from (11) and Lemma 8. □

The closed intervals in (17) are validated by the continuity Lemma 6, which further determines that

$$k_{\lfloor \overline{x} \rfloor} = k_0 f_{\tilde{N}}(\overline{x}) \left(\prod_{i=1}^{\lfloor \overline{x} \rfloor} f_{\tilde{N}}(\overline{x} - 2i) \right)^{-\frac{1}{\lfloor \overline{x} \rfloor}} \tag{19}$$

$$k_{\lfloor \overline{x} \rfloor + 1} = k_0 f_{\tilde{N}}(\overline{x}) \left(\prod_{i=0}^{\lfloor \overline{x} \rfloor} f_{\tilde{N}}(\overline{x} - 2i) \right)^{-\frac{1}{\lfloor \overline{x} \rfloor + 1}}. \tag{20}$$

The constraint (4) and Lemma 2 in turn determine that

$$k_0 = \frac{1}{2f_{\tilde{N}}(\overline{x})}\left\{(\lfloor\overline{x}\rfloor+1)\int_{\lfloor\overline{x}\rfloor}^{\overline{x}}\cdot\left(\prod_{i=0}^{\lfloor\overline{x}\rfloor}\frac{f_{\tilde{N}}(x-2i)}{f_{\tilde{N}}(\overline{x}-2i)}\right)^{\frac{1}{\lfloor\overline{x}\rfloor+1}}dx + \lfloor\overline{x}\rfloor\int_{\overline{x}-2}^{\lfloor\overline{x}\rfloor-1}\left(\prod_{i=0}^{\lfloor\overline{x}\rfloor-1}\frac{f_{\tilde{N}}(x-2i)}{f_{\tilde{N}}(\overline{x}-2-2i)}\right)^{\frac{1}{\lfloor\overline{x}\rfloor}}dx + \int_{\overline{x}}^{a}\frac{f_{\tilde{N}}(x)}{f_{\tilde{N}}(\overline{x})}\,dx\right\}^{-1}. \tag{21}$$

Straightforward computation yields for divergence the expression

$$\begin{aligned} D(f_{\hat{N}}\|f_N) &= D(f_{\hat{N}}\|f_{\tilde{N}}) + \log C - c\sigma^2 \\ &= 2A_0\log k_0 + 2\lfloor\overline{x}\rfloor A_{\lfloor\overline{x}\rfloor}\log k_{\lfloor\overline{x}\rfloor} \\ &\quad + 2(\lfloor\overline{x}\rfloor+1)A_{\lfloor\overline{x}\rfloor+1}\log k_{\lfloor\overline{x}\rfloor+1} + \log C - c\sigma^2 \\ A_0 &= \int_{\overline{x}}^{a} f_{\hat{N}}(x)\,dx \\ A_{\lfloor\overline{x}\rfloor} &= \int_{\lfloor\overline{x}\rfloor-1}^{2\lfloor\overline{x}\rfloor-\overline{x}} f_{\hat{N}}(x)\,dx \\ A_{\lfloor\overline{x}\rfloor+1} &= \int_{\lfloor\overline{x}\rfloor}^{\overline{x}} f_{\hat{N}}(x)\,dx \end{aligned} \tag{22}$$

and for the second moment of $f_{\hat{N}}$ the expression

$$\begin{aligned} E[\hat{N}^2] &= 2B_0 + 2B_{\lfloor\overline{x}\rfloor} + 2(\lfloor\overline{x}\rfloor+1)B_{\lfloor\overline{x}\rfloor+1} \\ &\quad + \frac{2(\lfloor\overline{x}\rfloor-1)\lfloor\overline{x}\rfloor(\lfloor\overline{x}\rfloor+1)}{3}A_{\lfloor\overline{x}\rfloor} \\ &\quad + \frac{2\lfloor\overline{x}\rfloor(\lfloor\overline{x}\rfloor+1)(\lfloor\overline{x}\rfloor+2)}{3}A_{\lfloor\overline{x}\rfloor+1} \\ B_0 &= \int_{\overline{x}}^{a} x^2 f_{\hat{N}}(x)\,dx \\ B_{\lfloor\overline{x}\rfloor} &= \int_{\lfloor\overline{x}\rfloor-1}^{2\lfloor\overline{x}\rfloor-\overline{x}} (x-(\lfloor\overline{x}\rfloor-1))^2 f_{\hat{N}}(x)\,dx \\ B_{\lfloor\overline{x}\rfloor+1} &= \int_{\lfloor\overline{x}\rfloor}^{\overline{x}} (x-\lfloor\overline{x}\rfloor)^2 f_{\hat{N}}(x)\,dx. \end{aligned} \tag{23}$$

Finally, the constants c and $\overline{x}$ are determined, respectively, by the relationships $D(f_{\hat{N}}\|f_N) = \delta$ via Lemma 4 and $E[\hat{N}^2] = \sigma^2$ via the following result.

Lemma 9: For those nominals satisfying the conditions of Theorem 2, the power constraint (5) is active in the optimization problem (3).

The worst case ML probability of detection error associated with $f_{\hat{N}}$ is given by

$$P_{\hat{N}} = \int_1^a f_{\hat{N}}(x)\,dx$$

according to Lemma 7. In the case

$$a \leq \lceil\sqrt{3\sigma^2+1} - 2$$

and $\delta > \delta_{N,\sigma^2}$, worst case noise falls in the class $\mathcal{M}(a)$ and exhibits

$$P_{\hat{N}} = \lfloor a\rfloor/(2(\lfloor a\rfloor+1)).$$

It is interesting to note that, according to Theorem 1, the worst case noise pdf governed by Theorem 2 should converge in distribution to the worst case power-constrained distribution $F_{\sigma^2}^{SV}$ as the divergence tolerance grows unbounded. It is clear from (17) that, as the constant c tends to infinity, all mass associated with $f_{\hat{N}}$ tends to the integers, directly implying convergence of the distribution $F_{\hat{N}}$ to $F_{\sigma^2}^{SV}$ by Lemma 9 and the fact that the unique symmetric 2-periodic probability mass function with second moment σ^2 and taking values on the integers is given by $\mathcal{P}_{\sigma^2}^{SV}$. While this suggests a straightforward approach for verification of the required convergence, it is unfortunately not immediately apparent how to establish the implication that c must tend to infinity as δ grows unbounded. Hence, we dismiss a rigorous verification in favor of this intuitively satisfying observation.

Theorem 2 and Lemma 7 together lead to the important observation that there is no saddle point in the two-person game between unconstrained receiver and noise constrained in power and divergence. This conclusion follows directly from the uniqueness of the worst case *zero-threshold* noise pdf determined in [18] and its discrepancy from the worst case pdf given by (17), for which the zero-threshold detector is an ML detector according to Lemma 7. Such a discrepancy demonstrates that when the maximin receiver is fixed, the maximin noise distribution can be perturbed within the constrained class to increase detection error probability, implying that a saddle point does not exist.

B. Divergence Constraint

For channels wherein the class of noise under consideration lies close to some prespecified distribution, but without necessarily exhibiting limited power, worst case ML divergence-constrained noise is the performance-characterizing distribution of interest. This problem corresponds exactly with the optimization problem (3) of Section II-A except for the dropping of the power constraint (5), leading to an enlarged feasible solution set and a somewhat simpler analysis.

The worst case divergence-constrained noise $\tilde{N}$ exhibits the pdf $f_{\tilde{N}}$ satisfying the optimization problem

$$P_{\tilde{N}} = \arg\max_{f_{\mathcal{N}}} P_{\mathrm{ML}}(f_{\mathcal{N}}) \tag{24}$$

subject to the constraints

$$\int_{-a}^{a} f_{\mathcal{N}}(x)\,dx = 1 \tag{25}$$

$$D(f_{\mathcal{N}}\|f_N) \leq \delta. \tag{26}$$

We denote by $P_{\tilde{N}}$ the achieved worst case ML probability of detection error $P_{\mathrm{ML}}(f_{\tilde{N}})$, and use the augmented notation $f_{\tilde{N}N,\delta}$ and $P_{\tilde{N},N,\delta}$ to represent $f_{\tilde{N}}$ and $P_{\tilde{N}}$ respectively whenever doing so makes for a clearer analysis.

By ignoring all aspects of the proofs of Lemmas 1, 2, and 5–7 associated with satisfaction of the power constraint (5), we obtain directly that each of these observations holds for a divergence-constrained analysis as well. Hence, we have the following:

Lemma 10: Given a continuous symmetric nominal pdf f_N with support $(-a, a)$, the solution $f_{\tilde{N}}$ to (24) is unique whenever the divergence constraint is active.

Lemma 11: Given a continuous symmetric nominal pdf f_N with support $(-a, a)$, the solution $f_{\tilde{N}}$ to (24) is symmetric whenever the divergence constraint is active.

Lemma 12: Any solution to the optimization problem obtained by replacing the divergence constraint (26) with the corresponding equality constraint

$$D(f_{\mathcal{N}} \| f_N) = \delta \tag{27}$$

is also a solution to (24).

Lemma 13: Given a continuous symmetric nominal pdf f_N with support $(-a, a)$, the solution $f_{\tilde{N}}$ to (24) is continuous whenever the divergence constraint is active.

Lemma 14: Given a symmetric unimodal nominal f_N, the zero-threshold detector is an ML detector for the worst case noise pdf $f_{\tilde{N}}$.

The previous treatment concerning characterization of activity of the divergence constraint in Lemma 4 does *not* carry over to the present analysis, primarily due to the general discrepancy in asymptotic behavior of worst case noise with and without a power constraint for large values of divergence tolerance. The proof of Lemma 4 makes use of the upper bound on worst case power- and divergence-constrained noise provided by the worst case power-constrained distribution $F_{\sigma^2}^{SV}$ developed in [13] as well as the worst case amplitude-constrained class $\mathcal{M}(a)$ developed in [21], depending on the relationship between the quantities a and $\lceil \sqrt{3\sigma^2 + 1} - 1 \rceil$. While worst case power-constrained error probability does not provide an upper bound on achieved error performance for the optimization problem (24), worst case amplitude-constrained error probability does so for nominals with finite support. For nominals with infinite support, for instance Gaussian distributions, neither quantity provides an upper bound, necessitating a separate analysis.

Lemma 15: Given any nominal f_N with finite support $(-a, a), a < \infty$, the divergence constraint (26) is active for all $\delta \leq \delta_N$ where

$$\delta_N = \min_{f_{\mathcal{N}} \in \mathcal{M}(a)} D(f_{\mathcal{N}} \| f_N).$$

The actual value of δ_N and asymptotic form of $f_{\tilde{N}, N, \delta}$ are given detailed attention in Section IV-E. For nominals with infinite support, we make the following more general observation concerning activity of the divergence constraint.

Lemma 16: Given a nominal f_N with infinite support, the divergence constraint (26) is active for all values of δ.

Analysis of the general form of $f_{\tilde{N}}$ is very similar to the support decomposition approach taken in Section II-A where the form of the solution $f_{\tilde{N}}$ subject to a power constraint was of interest. Again, we form the support decomposition

$$\Lambda = \{\mathcal{A}, \mathcal{B}, \overline{\mathcal{B}}, \mathcal{C}_2, \overline{\mathcal{C}}_2, \mathcal{C}_2^L, \mathcal{C}_2^R, \mathcal{C}_3, \overline{\mathcal{C}}_3, \mathcal{C}_3^L, \mathcal{C}_3^R, \cdots\}$$

as in Section II-A, and define the class $\mathcal{F}_\Lambda$ of pdf's exhibiting the support decomposition Λ. Then $f_{\tilde{N}}$ is the (not necessarily unique) pdf for which

$$P_{\mathrm{ML}}(f_{\tilde{N}}) = \max_{\Lambda} \max_{f_{\mathcal{N}} \in \mathcal{F}_\Lambda} P_{\mathrm{ML}}(f_{\mathcal{N}})$$

subject to the constraints (8) and (10) but not (9). Taking into account Lemma 12, a Lagrange-multiplier analysis yields for worst case noise $\tilde{N}$ the form

$$f_{\tilde{N}}(x) = \begin{cases} k_1 f_N(x), & x \in \mathcal{A} \\ k_0 f_N(x), & x \in \mathcal{B} \\ \overline{k}_0 f_N(x), & x \in \overline{\mathcal{B}} \\ k_n \left(\prod_{i=0}^{n-1} f_N(x-2i) \right)^{\frac{1}{n}}, & \\ \qquad x \in \mathcal{C}_n, n = 2, 3, \cdots & \\ \overline{k_n} \left(\prod_{i=0}^{n-1} f_N(x-2i) \right)^{\frac{1}{n}}, & \\ \qquad x \in \overline{\mathcal{C}}_n, n = 2, 3, \cdots & \\ k_n^* \left(\prod_{i=0}^{n-1} f_N(x-2i) \right)^{\frac{1}{n}}, & \\ \qquad x \in \mathcal{C}_n^L \cup \mathcal{C}_n^R, n = 2, 3, \cdots & \end{cases} \tag{28}$$

where $k_0, \overline{k}_0, k_1, k_2, \overline{k}_2, k_2^*, k_3, \overline{k}_3, k_3^*, \cdots$ are constants.

The form of support decomposition Λ can again be simplified using Lemma 8 for the class of symmetric continuous unimodal nominals satisfying condition (15), leading to the following result.

Theorem 3: Given a continuous symmetric unimodal nominal f_N with support $(-a, a)$ and satisfying (15), the solution to (24) is given by

$$f_{\tilde{N}}(x) = \begin{cases} k_0 f_N(x), & |x| > \overline{x} \\ k_{\lfloor \overline{x} \rfloor} \left(\prod_{i=0}^{\lfloor \overline{x} \rfloor - 1} f_N(x-2i) \right)^{\frac{1}{\lfloor \overline{x} \rfloor}}, & \\ \qquad x \in [\overline{x} - 2, 2\lfloor \overline{x} \rfloor - \overline{x}] & \\ k_{\lfloor \overline{x} \rfloor + 1} \left(\prod_{i=0}^{\lfloor \overline{x} \rfloor} f_N(x-2i) \right)^{\frac{1}{\lfloor \overline{x} \rfloor + 1}}, & \\ \qquad x \in [2\lfloor \overline{x} \rfloor - \overline{x}, \overline{x}] & \\ f_{\tilde{N}}(x+2i), & \\ \qquad x + 2i \in [\overline{x} - 2, 2\lfloor \overline{x} \rfloor - \overline{x}], i = 1, 2, \cdots, \lfloor \overline{x} \rfloor - 1 & \\ f_{\tilde{N}}(x+2i), & \\ \qquad x + 2i \in [2[\overline{x}] - \overline{x}, \overline{x}], i = 1, 2, \cdots, \lfloor \overline{x} \rfloor & \end{cases} \tag{29}$$

where $\overline{x} \in (1, a)$.

Proof: The result follows directly from (28) and Lemma 8. □

Care must be taken in characterizing activity of the divergence constraint according to Lemmas 15 and 16, which is of interest in order to apply Lemma 13 to reduce the number of degrees of freedom associated with the weighting constants in (28) and (29). If the nominal f_N has infinite support, the divergence constraint is active, as is the case if f_N has support $(-a, a), a < \infty$ and $\delta \le \delta_N$ where

$$\delta_N = \min_{f_{\mathcal{N}} \in \mathcal{M}(a)} D(f_{\mathcal{N}} \| f_N)$$

a quantity whose evaluation is discussed in Section IV-E. In either of these cases, the continuity Lemma 13 determines that

$$k_{\lfloor \overline{x} \rfloor} = k_0 f_N(\overline{x}) \left(\prod_{i=1}^{\lfloor \overline{x} \rfloor} f_N(\overline{x} - 2i) \right)^{-\frac{1}{\lfloor \overline{x} \rfloor}}$$

and

$$k_{\lfloor \overline{x} \rfloor + 1} = k_0 f_N(\overline{x}) \left(\prod_{i=0}^{\lfloor \overline{x} \rfloor} f_N(\overline{x} - 2i) \right)^{-\frac{1}{\lfloor \overline{x} \rfloor + 1}}$$

which, along with the fact that $f_{\hat{N}}$ is a proper pdf, determine that

$$\begin{aligned} k_0 = \frac{1}{2 f_N(\overline{x})} \Bigg\{ (\lfloor \overline{x} \rfloor + 1) \int_{\lfloor \overline{x} \rfloor}^{\overline{x}} & \\ \cdot \left(\prod_{i=0}^{\lfloor \overline{x} \rfloor} \frac{f_N(x - 2i)}{f_N(\overline{x} - 2i)} \right)^{\frac{1}{\lfloor \overline{x} \rfloor + 1}} dx & \\ + \lfloor \overline{x} \rfloor \int_{\overline{x} - 2}^{\lfloor \overline{x} \rfloor - 1} \left(\prod_{i=0}^{\lfloor \overline{x} \rfloor - 1} \frac{f_N(x - 2i)}{f_N(\overline{x} - 2 - 2i)} \right)^{\frac{1}{\lfloor \overline{x} \rfloor}} dx & \\ + \int_{\overline{x}}^{a} \frac{f_N(x)}{f_N(\overline{x})} \, dx \Bigg\}^{-1}. & \end{aligned}$$

Finally, $\overline{x}$ is determined in the case of an active divergence constraint by setting

$$\int_{-a}^{a} f_{\hat{N}}(x) \log \frac{f_{\hat{N}}(x)}{f_N(x)} \, dx = \delta.$$

C. Nominals with Discontinuities

The majority of results developed thus far assume that the nominal noise pdf is continuous symmetric with support over a single interval $(-a, a), 1 < a \le \infty$. The relaxation of the support assumption through transportation of negligible quantities of mass was discussed in Section II-A. Here we discuss the relaxation of the continuity assumption by introducing a simple procedure for treating nominals with discontinuities.

Given a bounded symmetric nominal pdf f_N with support over an interval $(-a, a)$ (where a need not be finite) and a countable number of discontinuities, consider any sequence $\{f_{N,i}\}_{i=1}^{\infty}$ of bounded *continuous* pdf's satisfying $\lim_{i \to \infty} f_{N,i} = f_N$, each exhibiting a bounded worst case power- and divergence-constrained pdf $f_{\hat{N},i}$ determined by the analysis of Section II-A. Define $f_L = \lim_{i \to \infty} f_{\hat{N},i}$ and note through bounded convergence that

$$\begin{aligned} \lim_{i \to \infty} \int_{-a}^{a} f_{\hat{N},i}(x) \, dx &= \int_{-a}^{a} f_L(x) \, dx \\ \lim_{i \to \infty} \int_{-a}^{a} x^2 f_{\hat{N},i}(x) \, dx &= \int_{-a}^{a} x^2 f_L(x) \, dx \\ \lim_{i \to \infty} D(f_{\hat{N},i} \| f_N) &= D(f_L \| f_N) \\ \lim_{i \to \infty} P_{\mathrm{ML}}(f_{\hat{N},i}) &= P_{\mathrm{ML}}(f_L) \end{aligned}$$

yielding the desired result $f_{\hat{N}} \equiv f_L$. Coupled with the strategy for treating nominals with support over disjoint intervals discussed above, this procedure allows for application of the results of Section II-A to any symmetric nominal pdf. This procedure clearly extends to the case of unconstrained power as well, allowing for the application of the results of Section II-B to any symmetric nominal pdf.

III. Gaussian Nominals

An analysis of near-Gaussian channels is performed by specifying a Gaussian pdf for the nominal centering the divergence-constrained candidate class of worst case noise distributions. Such a specification encompasses models incorporating an ambient Gaussian noise channel subject to interference, jamming, or any general contamination satisfying a proximity (divergence) bound. In this case, the divergence-from-Gaussian constraint provides a measure of "non-Gaussianness" as defined in [17] and [22].

A. Power- and Divergence-from-Gaussian Constraints

Assume the nominal noise N is a zero-mean, variance-σ^2 Gaussian random variable, denoted by $\mathcal{N}(0, \sigma^2)$. The nominal pdf $f_{\mathcal{N}(0,\sigma^2)}(x)$ then takes the form

$$f_{\mathcal{N}(0,\sigma^2)}(x) = \frac{1}{\sigma} \, \Phi\left(\frac{x}{\sigma}\right)$$

where

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \, e^{-\frac{x^2}{2}}$$

is the standard Gaussian pdf. Given this prespecification, worst case power- and divergence-constrained noise $f_{\hat{N}}$ takes the form (17) according to Theorem 2. Direct substitution in (18) yields

$$f_{\hat{N}}(x) = \frac{\sqrt{1 + 2c\sigma^2}}{\sqrt{2\pi}\sigma} \, e^{-\frac{x^2(1+2c\sigma^2)}{2\sigma^2}}.$$

Note that *the variance-scaled nominal is a Gaussian pdf whenever the nominal is Gaussian,* with second moment given by $\overline{\sigma}^2 = \sigma^2(1 + 2c\sigma^2)^{-1}$. Direct substitution also shows that the geometric mean appearing in (17) takes the form

$$\left(\prod_{i=0}^{\lfloor \overline{x} \rfloor} f_{\bar{N}}(x - 2i) \right)^{\frac{1}{\lfloor \overline{x} \rfloor + 1}} = \frac{e^{-\frac{\lfloor \overline{x} \rfloor (\lfloor \overline{x} \rfloor + 2)}{6\overline{\sigma}^2}}}{\sqrt{2\pi}\overline{\sigma}} \, e^{-\frac{(x - \lfloor \overline{x} \rfloor)^2}{2\overline{\sigma}^2}}$$

corresponding to a weighted $\mathcal{N}(\lfloor \overline{x} \rfloor, \overline{\sigma}^2)$ distribution. This demonstrates that the worst case pdf is made up of a mixture of truncated variance-$\overline{\sigma}^2$ Gaussians. For fixed c and $\overline{x}$, the solution takes the form

$$f_{\hat{N}}(x) = \begin{cases} \dfrac{k_0}{\sqrt{2\pi}\overline{\sigma}}\, e^{-\frac{x^2}{2\overline{\sigma}^2}}, \qquad |x| > \overline{x} \\ \dfrac{k_{\lfloor\overline{x}\rfloor}\, e^{-\frac{(\lfloor\overline{x}\rfloor-1)(\lfloor\overline{x}\rfloor+1)}{6\overline{\sigma}^2}}}{\sqrt{2\pi}\overline{\sigma}} e^{-\frac{(x-(\lfloor\overline{x}\rfloor-1))^2}{2\overline{\sigma}^2}}, \\ \qquad\qquad x \in [\overline{x}-2, 2\lfloor\overline{x}\rfloor-\overline{x}] \\ \dfrac{k_{\lfloor\overline{x}\rfloor+1}\, e^{-\frac{\lfloor\overline{x}\rfloor(\lfloor\overline{x}\rfloor+2)}{6\overline{\sigma}^2}}}{\sqrt{2\pi}\overline{\sigma}}\, e^{-\frac{(x-\lfloor\overline{x}\rfloor)^2}{2\overline{\sigma}^2)}}, \quad x \in [\overline{x}-2, \overline{x}] \\ f_{\hat{N}}(x+2i), \\ \qquad x+2i \in [\overline{x}-2, 2\lfloor\overline{x}\rfloor-\overline{x}], i = 1, 2, \cdots, \\ \qquad\qquad \lfloor\overline{x}\rfloor - 1 \\ f_{\hat{N}}(x+2i), \\ \qquad x+2i \in [2[\overline{x}]-\overline{x}, \overline{x}], i = 1, 2, \cdots, \lfloor\overline{x}\rfloor. \end{cases} \tag{30}$$

The constants $k_0, k_{\lfloor\overline{x}\rfloor}$ and $k_{\lfloor\overline{x}\rfloor+1}$ are given through (19)–(21) by

$$\begin{aligned} k_0 = \frac{1}{2} \Bigg\{ &(\lfloor\overline{x}\rfloor+1)e^{-\frac{\lfloor\overline{x}\rfloor(\lfloor\overline{x}\rfloor+2\tilde{x})}{2\overline{\sigma}^2}} \left(\frac{1}{2} - Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right) \\ &+ \lfloor\overline{x}\rfloor\, e^{-\frac{(\lfloor\overline{x}\rfloor+1)(\lfloor\overline{x}\rfloor-1+2\tilde{x})}{2\overline{\sigma}^2}} \\ &\cdot \left(\frac{1}{2} - Q\left(\frac{1-\tilde{x}}{\overline{\sigma}}\right)\right) + Q\left(\frac{\overline{x}}{\overline{\sigma}}\right)\Bigg\}^{-1} \end{aligned} \tag{31}$$

$$k_{\lfloor\overline{x}\rfloor} = k_0 e^{-\frac{(\lfloor\overline{x}\rfloor+1)(\lfloor\overline{x}\rfloor-1+3\tilde{x})}{3\overline{\sigma}^2}} \tag{32}$$

$$k_{\lfloor\overline{x}\rfloor+1} = k_0 e^{-\frac{\lfloor\overline{x}\rfloor(\lfloor\overline{x}\rfloor-1+3\tilde{x})}{3\overline{\sigma}}} \tag{33}$$

where $\tilde{x} = \overline{x} - \lfloor\overline{x}\rfloor$ and $Q(\cdot)$ represents the complimentary unit-variance Gaussian cumulative distribution function. The divergence and second moment of $f_{\hat{N}}$ are given by the expressions

$$\begin{aligned} D(f_{\hat{N}} \| \mathcal{N}(0,\sigma^2)) &= 2k_0 \log k_0 Q\left(\frac{\overline{x}}{\overline{\sigma}}\right) + 2\lfloor\overline{x}\rfloor k_{\lfloor\overline{x}\rfloor} \log k_{\lfloor\overline{x}\rfloor} \\ &\quad \cdot e^{-\frac{(\lfloor\overline{x}\rfloor-1)(\lfloor\overline{x}\rfloor+1)}{6\overline{\sigma}^2}} \\ &\quad \cdot \left[\frac{1}{2} - Q\left(\frac{1-\tilde{x}}{\overline{\sigma}}\right)\right] \\ &\quad + 2(\lfloor\overline{x}\rfloor+1)k_{\lfloor\overline{x}\rfloor+1} \log k_{\lfloor\overline{x}\rfloor+1} \\ &\quad \cdot e^{-\frac{\lfloor\overline{x}\rfloor(\lfloor\overline{x}\rfloor+2)}{6\overline{\sigma}^2}} \left[\frac{1}{2} - Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right] \\ &\quad + \frac{1}{2}\log(1+2c\sigma^2) - c\sigma^2 \end{aligned} \tag{34}$$

$$\begin{aligned} E[\hat{N}^2] = 2k_0 &\left[\frac{\overline{\sigma x}}{\sqrt{2\pi}}\, e^{\frac{\lfloor\overline{x}\rfloor}{2\overline{\sigma}^2}} + \overline{\sigma}^2 Q\left(\frac{\overline{\sigma}^2}{\overline{\sigma}}\right)\right] \\ &+ 2\lfloor\overline{x}\rfloor k_{\lfloor\overline{x}\rfloor}\, e^{-\frac{(\lfloor\overline{x}\rfloor-1)(\lfloor\overline{x}\rfloor+1)}{6\overline{\sigma}^2}} \\ &\cdot \Bigg[\left(\overline{\sigma}^2 + \frac{(\lfloor\overline{x}\rfloor-1)(\lfloor\overline{x}\rfloor+1)}{3}\right) \\ &\quad \times \left(\frac{1}{2} - Q\left(\frac{1-\tilde{x}}{\overline{\sigma}}\right)\right) \\ &\quad - \frac{\overline{\sigma}(1-\tilde{x})}{\sqrt{2\pi}}\, e^{-\frac{(1-\tilde{x})^2}{2\overline{\sigma}^2}}\Bigg] \\ &+ 2(\lfloor\overline{x}\rfloor+1)k_{\lfloor\overline{x}\rfloor+1}\, e^{-\frac{\lfloor\overline{x}\rfloor(\lfloor\overline{x}\rfloor+2)}{6\overline{\sigma}^2}} \\ &\cdot \Bigg[\left(\overline{\sigma}^2 + \frac{\lfloor\overline{x}\rfloor(\lfloor\overline{x}\rfloor+2)}{3}\right) \\ &\quad \times \left(\frac{1}{2} - Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right) - \frac{\overline{\sigma}\tilde{x}}{\sqrt{2\pi}}\, e^{-\frac{\tilde{x}^2}{2\overline{\sigma}^2}}\Bigg] \end{aligned} \tag{35}$$

through (22) and (23), making use of the identity

$$\int_z^\infty x^2\Phi(x)\,dx = \frac{z}{\sqrt{2\pi}}\, e^{-\frac{z^2}{2}} + Q(z)$$

for Gaussian pdf's. Finally, the worst case probability of ML detection error is given by

$$P_{\hat{N}} = \int_1^\infty f_{\hat{N}}(x)\,dx. \tag{36}$$

Note that the expression (30) describes a worst case pdf made up of a 2-periodic interval $[-\overline{x}, \overline{x}]$ with $\mathcal{N}(0, \overline{\sigma}^2)$ tails. The interval $[-\overline{x}, \overline{x}]$ consists of a symmetric mixture of truncated variance-$\overline{\sigma}^2$ Gaussians; each of the points $-\lfloor\overline{x}\rfloor, -\lfloor\overline{x}\rfloor + 2, \cdots, \lfloor\overline{x}\rfloor$ centers a weighted variance-$\overline{\sigma}^2$ Gaussian pdf truncated to an interval of length $2\tilde{x}$, while each of the points $-\lfloor\overline{x}\rfloor+1, -\lfloor\overline{x}\rfloor+3, \cdots, \lfloor\overline{x}\rfloor-1$ centers a similar Gaussian pdf truncated to an interval of length $2-2\tilde{x}$. The two degrees of freedom corresponding to the parametric constants c and $\overline{x}$ are determined by the relations $D(f_{\hat{N}} \| \mathcal{N}(0,\sigma^2)) = \delta$ and $E[\hat{N}^2] = \sigma^2$ through (34) and (35).

B. Divergence-from-Gaussian Constraint

The divergence-constrained problem for Gaussian nominals is governed by Theorem 3, which by direct substitution yields the form (30) for worst case noise pdf with the simple modification that $\overline{\sigma}^2 = \sigma^2$. Following the analysis of Section III-A and making the appropriate modification, worst case divergence-constrained noise for a Gaussian nominal is made up of a 2-periodic interval $[-\overline{x}, \overline{x}]$ with zero-mean, variance-σ^2 Gaussian tails. The interval $[-\overline{x}, \overline{x}]$ is made up of a mixture of truncated variance-σ^2 Gaussians; each of the points $-\lfloor\overline{x}\rfloor, -\lfloor\overline{x}\rfloor+2, \cdots, \lfloor\overline{x}\rfloor$ centers a weighted variance-σ^2 Gaussian pdf truncated to an interval of length $2\tilde{x}$ where $\tilde{x} = \overline{x} - \lfloor\overline{x}\rfloor$, while each of the points $-\lfloor\overline{x}\rfloor+1, \lfloor\overline{x}\rfloor+3, \cdots, \lfloor\overline{x}\rfloor-1$ centers a similar Gaussian truncated to an interval of length $2-2\tilde{x}$.

Fig. 3 depicts worst case noise for a nominal Gaussian channel with SNR $=$ 10 dB for a variety of divergence tolerance values, with and without a power constraint. Fig. 4 depicts similar curves for a channel with SNR $=$ -8.45 dB. All peaks are piecewise weighted Gaussians. According to Theorem 1, the power- and divergence-constrained pdf's of Fig. 3 tend, as δ grows, to a three-point mass function with weights 0.05 at the points $x = -1, 1$ and weight 0.9 at $x = 0$ as determined in [13]. Similarly, the limiting distribution of the power- and divergence-constrained pdf's of Fig. 4 is given in [13, p.1502]. The worst case pdf's without a power constraint are made up of a mixture of piecewise weighted translations of the nominal Gaussian, with ML error probability tending to the upper bound of $1/2$ with growing divergence tolerance according to the analysis of

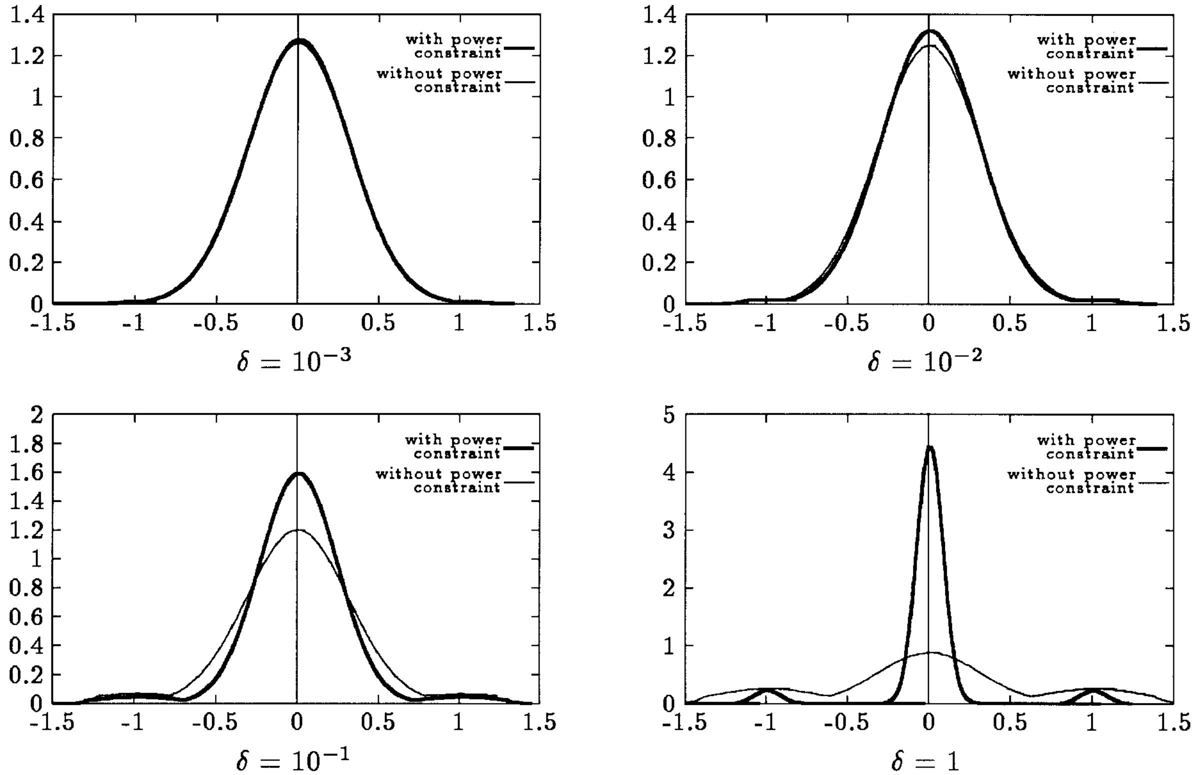

Fig. 3. Worst case ML noise pdf's for a Gaussian nominal, SNR $= 10$ dB.

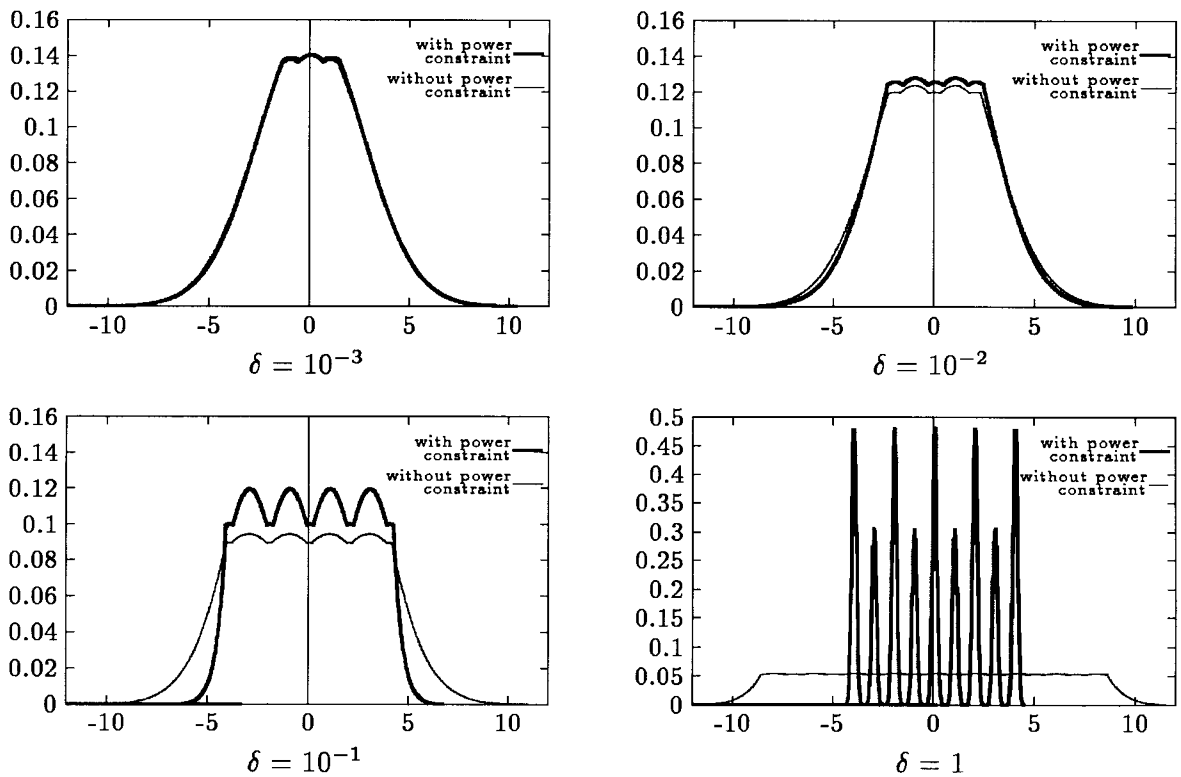

Fig. 4. Worst case ML noise pdf's for a Gaussian nominal, SNR $= -8.45$ dB.

Section IV-E. Of note is the increase in discrepancy between the worst case pdf's with and without a power constraint for growing divergence tolerance, an expected feature of the enlarged feasible classes.

IV. Asymptotic Behavior

A. Small Divergence Tolerance with Power Constraint

We are interested in studying the asymptotic behavior of worst case power- and divergence-constrained noise for Gaussian nominals as the divergence tolerance tends to zero. Noting that the divergence between any two pdf's is zero if and only if they are identical (up to measure zero), this represents the scenario where a fixed-power noise class exhibits near-Gaussian behavior. For instance, the performance of a multiuser communication system wherein the superposition of multiple-access interference and background noise displays quasi-Gaussian characteristics warrants such an analysis.

Proposition 1: Given a $\mathcal{N}(0,\sigma^2)$ nominal pdf and fixed SNR, the worst case power- and divergence-constrained error degradation satisfies the asymptotic relation

$$\lim_{\delta\to 0}\frac{P_{\hat{N},\mathcal{N}(0,\sigma^2),\delta}-Q_{\sigma^2}}{\sqrt{\delta}}=\sqrt{Q_{\sigma^2}(1-2Q_{\sigma^2})-\frac{e^{-\frac{1}{\sigma^2}}}{2\pi\sigma^2}}$$

where $Q_{\sigma^2}=Q(1/\sigma)$ is the nominal ML error probability.

For a derivation, see the Appendix.

For high SNR channels, we see that

$$P_{\hat{N},\mathcal{N}(0,\sigma^2),\delta}\simeq Q_{\sigma^2}+\sqrt{Q_{\sigma^2}}\sqrt{\delta}+o(\sqrt{\delta})$$

demonstrating that worst case error performance is sensitive to non-Gaussianness even when the receiver is allowed to optimally adjust to the noise distribution.

B. Small Divergence Tolerance Without Power Constraint

Consider worst case divergence-constrained noise for a Gaussian nominal as the divergence tolerance tends to zero. Such a scenario arises when, for instance, a decision is warranted as to whether a given unknown channel fits the standard AWGN model. This decision is discussed as being reliably achievable in [23] given a large enough number of transmitted symbols and a stationarity assumption on the noise, where the associated hypothesis test is based on a divergence threshold δ_0; given the hypothesis corresponding to a Gaussian fit, this threshold provides the appropriate divergence bound for (24), making the present results of immediate interest in quantifying guaranteed channel performance.

Proposition 2: Given a $\mathcal{N}(0,\sigma^2)$ nominal pdf and fixed SNR, the worst case divergence-constrained error degradation satisfies the following asymptotic relation (derived in the Appendix):

$$\lim_{\delta\to 0}\frac{P_{\tilde{N},\mathcal{N}(0,\sigma^2),\delta}-Q_{\sigma^2}}{\sqrt{\delta}}=\sqrt{Q_{\sigma^2}(1-2Q_{\sigma^2})}.$$

This result provides for an interesting comparison with the expression

$$\lim_{\delta\to 0}\frac{\Delta P_T}{\sqrt{\delta}}=\sqrt{Q_{\sigma^2}(1-2Q_{\sigma^2})}\tag{37}$$

found in [18] governing asymptotic behavior of worst case divergence-constrained error degradation for a zero-threshold detector. The implication is that, given a noise class falling in a very small divergence neighborhood of a Gaussian distribution (i.e., a quasi-Gaussian channel), there is no loss in worst case performance incurred by using a hard-limiter in place of an ML decision rule.

C. High Signal-to-Noise Ratio

We turn our attention to the analysis of worst case power- and divergence-constrained noise for Gaussian nominals as the noise power tends to zero. This represents a channel subject to noise that may stray significantly from a nominal Gaussian distribution, but is known to exhibit high SNR. The results pertain to reliable transmission applications where noise is non-Gaussian.

Proposition 3: Given the class of $\mathcal{N}(0,\sigma^2)$ nominal pdf's parameterized by σ^2 and fixed divergence tolerance δ, the worst case power- and divergence-constrained ML error probability satisfies the following asymptotic relation (derived in the Appendix):

$$\lim_{\sigma^2\to 0}\frac{P_{\hat{N},\mathcal{N}(0,\sigma^2),\delta}}{\sigma^2}=\frac{1-e^{-2\delta}}{4}.$$

Such behavior is demonstrated in Fig. 1 where the divergence-constrained curve generated according to Theorem 2 exhibits a near-linear dependence on σ^2 for high SNR values, with the ratio given approximately by $(1-e^{-2\delta})/4\simeq\delta/2$ for small values of δ. An informative manner in which to view this result derives from a comparison with worst case power-constrained noise for ML detection and asymptotically high SNR, shown in [13] to take the form of a probability mass function $\mathcal{P}_{\sigma^2}^{SV}$ taking the value $\sigma^2/2$ at the points ± 1 and the value $1-\sigma^2$ at zero, achieving probability of ML detection error $P_{SV}=\sigma^2/4$. Noting that P_{SV} corresponds with $P_{\hat{N},\mathcal{N}(0,\sigma^2),\delta}$ when $\delta=\infty$, we have the result

$$\lim_{\sigma^2\to 0}\frac{P_{\hat{N},\mathcal{N}(0,\sigma^2),\delta}}{P_{\hat{N},\mathcal{N}(0,\sigma^2),\infty}}=1-e^{-2\delta}.\tag{38}$$

Hence, the imposition of a divergence-from-Gaussian constraint results in a linear reduction in worst case power-constrained error performance, the ratio being given by $1-e^{-2\delta}$ where δ is the imposed divergence tolerance.

D. Large Divergence Tolerance with Power Constraint

In this section we analyze the behavior of worst case power- and divergence-constrained noise as the divergence tolerance is allowed to grow unbounded, representing the scenario where the power constraint dominates the optimization problem (3).

Proposition 4: Given a nominal pdf with second moment σ^2 and support $(-a, a)$, the worst case power- and divergence-constrained ML detection error probability satisfies

$$\lim_{\delta \to \infty} P_{\hat{N},N,\delta} = \frac{2K^3 - 2K + 3\sigma^2}{2K(K+1)(2K+1)},$$
$$K = \min\left\{\text{integers } k\text{: } \frac{k(k+2)}{3} \geq \sigma^2\right\}$$

if $a > \lceil \sqrt{3\sigma^2 + 1} - 1 \rceil$ and

$$P_{\hat{N},N,\delta} = \frac{\lceil a \rceil}{2(\lceil a \rceil + 1)}$$

for all $\delta \geq \delta_{N,\sigma^2}$ if

$$a \leq \lceil \sqrt{3\sigma^2 + 1} - 1 \rceil.$$

For a derivation and additional asymptotic analysis, including an expression for δ_{N,σ^2}, see the Appendix.

E. Large Divergence Tolerance Without Power Constraint

Channels subject to a large class of potential noise distributions centered about some nominal can be studied by characterizing the behavior of worst case divergence-constrained noise for large values of divergence tolerance. This corresponds to the scenario where the noise distribution is derived from a prespecified nominal by some perturbation which can significantly affect the noise characteristics.

Proposition 5: Given a nominal pdf f_N with finite support $(-a, a)$, the worst case divergence-constrained ML detection error probability satisfies

$$P_{\tilde{N},N,\delta} = \frac{\lfloor a \rfloor}{2(\lfloor a \rfloor + 1)}$$

for all $\delta \geq \delta_N$. Given a nominal f_N with infinite support

$$\lim_{\delta \to \infty} P_{\tilde{N},N,\delta} = \tfrac{1}{2}.$$

For a derivation and additional asymptotic analysis, including an expression for δ_N, see the Appendix.

V. Conclusions

In Section I, a class of noise pdf was defined and motivated for binary-input channels based on constraints on power and divergence from a nominal noise distribution. Section II formed a characterization of worst case noise within this class, using minimum detection error probability as the optimality criterion. Worst case noise subject to a lone divergence constraint was also characterized.

In Section II-A, a unique worst case pdf $f_{\hat{N}}$ was shown to exist under an active divergence constraint based on assumptions of continuity, symmetry, and single-interval support $(-a, a)$ for the nominal f_N. Under such conditions, $f_{\hat{N}}$ was also shown to be symmetric and continuous. The limiting behavior of $f_{\hat{N}}$ and corresponding worst case probability of ML detection error $P_{\hat{N}}$ were characterized for asymptotically large values of divergence tolerance; this behavior was shown to be governed by the worst case power-constrained solution developed in [13] for nominals satisfying $a > \lceil \sqrt{3\sigma^2 + 1} - 1 \rceil$ where σ^2 is the power associated with f_N, and by the worst case amplitude-constrained solution developed in [21] for the case $a \leq \lceil \sqrt{3\sigma^2 + 1} - 1 \rceil$. Based on this analysis, activity of the divergence constraint was also characterized. For unimodal nominals, it was shown that the zero-threshold detector is an ML detector for $f_{\hat{N}}$, a useful property for both qualitative and quantitative analysis. A piecewise characterization of $f_{\hat{N}}$ was developed (11) along with a full expression (17) for nominals satisfying a weak local log-concavity condition (15), a condition satisfied by strictly log-concave pdf's (for example, Gaussians). In this case, the worst case pdf takes the form of a 2-periodic function over an interval $[-\overline{x}, \overline{x}], \overline{x} > 1$ made up of piecewise weighted geometric means derived from the variance-scaled nominal pdf $f_{\bar{N}}(x) = Cf_N(x)e^{-cx^2}$, while the tails are simply weighted tails of $f_{\bar{N}}$. Through the demonstrated continuity of $f_{\hat{N}}$, its parametric expression reduces to two degrees of freedom determined by activity of the power and divergence constraints.

Worst case divergence-constrained noise was analyzed in Section II-B upon dropping the power constraint. Results developed for the power- and divergence-constrained problem concerning uniqueness, symmetry, and continuity were shown to hold, along with results demonstrating ML performance of the zero-threshold detector. An expression for the worst case divergence-constrained noise pdf $f_{\tilde{N}}$ was developed (29) for nominals satisfying condition (15), displaying the same form as worst case power- and divergence-constrained noise upon substitution of the nominal f_N for the variance-scaled nominal $f_{\bar{N}}$.

In Section II-C, we lifted the technical assumptions on continuity and support, thereby extending results to symmetric nominal pdfs.

The important particular class of Gaussian nominals was analyzed in detail in Section III. An expression for worst case power- and divergence-from-Gaussian-constrained noise was developed (30) along with an expression for worst case probability of ML detection error (36). The scenario without power constraint was analyzed in Section III-B. It was found that the form of worst case noise is significantly simplified in this special case.

In Section IV, the behaviors of worst case noise and worst case probability of ML detection error were characterized for a variety of asymptotic scenarios. For Gaussian nominals subject to fixed power and asymptotically small divergence tolerance, the worst case error probability $P_{\hat{N},\mathcal{N}(0,\sigma^2),\delta}$ was shown to satisfy

$$P_{\hat{N},\mathcal{N}(0,\sigma^2),\delta} = Q_{\sigma^2} + \sqrt{Q_{\sigma^2}(1 - 2Q_{\sigma^2}) - \frac{e^{-\frac{1}{\sigma^2}}}{2\pi\sigma^2}} \cdot \sqrt{\delta} + o(\sqrt{\delta})$$

where $Q_{\sigma^2} = Q(1/\sigma)$ is the nominal probability of ML detection error. We conclude that worst case performance is very sensitive to non-Gaussianness of channel noise (as measured by divergence-from-Gaussian) even when the receiver is allowed to optimally adjust to the noise distribution. The worst case error probability $P_{\tilde{N},\mathcal{N}(0,\sigma^2),\delta}$ for the divergence-constrained problem with no power constraint was shown to satisfy

$$P_{\tilde{N},\mathcal{N}(0,\sigma^2),\delta} = Q_{\sigma^2} + \sqrt{Q_{\sigma^2}(1-2Q_{\sigma^2})}\,\sqrt{\delta} + o(\sqrt{\delta})$$

which coincides with the corresponding expression obtained for zero-threshold detection in [18], allowing us to conclude that there is no asymptotic loss incurred by employment of fixed zero-threshold detection rather than optimal detection for channels exhibiting arbitrarily small non-Gaussianness.

Worst case performance for Gaussian nominals, asymptotically high SNR and a fixed divergence tolerance δ was shown to satisfy

$$P_{\hat{N},\mathcal{N}(0,\sigma^2),\delta} = \frac{1-e^{-2\delta}}{4}\sigma^2 + o(\sigma^2).$$

We conclude that near–far resistance cannot be guaranteed in multiuser channels wherein pre-threshold processing leads to a noise distribution that can only be bounded in power and non-Gaussianness.

The behavior of worst case noise was characterized for growing divergence tolerance, including the development of expressions for the range of divergence tolerances leading to an active divergence constraint both with and without a power constraint.

Finally, it was demonstrated that, for symmetric unimodal nominals satisfying (15), there is no saddle point in the two-person game of unconstrained receiver versus power- and divergence-constrained noise.

Appendix
Proofs and Derivations

A. Proof of Lemma 1

Assume by contradiction that there exist two distinct solutions f_1 and f_2 for (3). Since the divergence constraint is assumed to be active, $D(f_1\|f_N) = D(f_2\|f_N) = \delta$. Consider the candidate pdf $f_3(x) = (f_1(x)+f_2(x))/2$. By convexity of the feasible set, f_3 is a feasible candidate, and performs as least as well as f_1 and f_2 by the noted concavity of the objective function (2). By strict convexity of the divergence measure and the assumption that f_1 and f_2 are distinct, we have that $D(f_3\|f_N) < \delta$, which leads through the assumed activity of the divergence constraint to the desired contradiction. □

B. Proof of Lemma 2

Assume by contradiction that the unique solution $f_{\hat{N}}$ is not symmetric. Note that

$$P_{\rm ML}(f_{\hat{N}}(x)) = P_{\rm ML}(f_{\hat{N}}(-x))$$

and that

$$D(f_{\hat{N}}(x)\|f_N(x)) = D(f_{\hat{N}}(-x)\|f_N(x)) = \delta$$

by the assumption that f_N is symmetric. Define the candidate

$$f'_{\hat{N}}p(x) = (f_{\hat{N}}(x) + f_{\hat{N}}(-x))/2.$$

By convexity of the feasible set and concavity of the objective function, $f'_{\hat{N}}$ is a feasible solution which performs at least as well as $f_{\hat{N}}$. The strict convexity of the divergence measure and the assumed asymmetry of $f_{\hat{N}}$ dictate that $D(f'_{\hat{N}}\|f_N) < \delta$, which leads to the desired contradiction in light of the assumed activity of the divergence constraint. □

C. Proof of Theorem 1

Choose any suitable nominal pdf f_N. An upper bound on the probability of ML detection error achieved by $f_{\hat{N},N,\delta}$ for any finite divergence tolerance δ is immediately provided by that associated with $\mathcal{P}^{SV}_{\sigma^2}$ since the latter corresponds exactly with the former except for the dropping of the divergence constraint. Our goal then is to describe a construction $f_{\varepsilon,N,\delta}$ for large δ among the valid class of candidates for (3) whose probability of ML detection error can be forced arbitrarily close to that of $\mathcal{P}^{SV}_{\sigma^2}$ by taking δ sufficiently large; by the uniqueness of $\mathcal{P}^{SV}_{\sigma^2}$ as a solution to the power-constrained problem (without divergence constraint), it directly follows that

$$\lim_{\delta\to\infty} F_{\hat{N},N,\delta} = F^{SV}_{\sigma^2}.$$

By supposition, there exists $\varepsilon > 0$ small enough so that the support of f_N includes the closed interval $[-M_{\sigma^2}-\varepsilon, M_{\sigma^2}+\varepsilon]$. There also exists $\varepsilon > 0$ small enough so that the unit-mass pulse

$$\Pi_\varepsilon(x) = \frac{1}{2\varepsilon}\,(u(x+\varepsilon) - u(x-\varepsilon))$$

exhibits second moment less than σ^2, where $u(\cdot)$ represents the unit step function

$$u(x) = \begin{cases} 0, & x < 0 \\ 1 & x \geq 0. \end{cases}$$

Take $0 < \varepsilon < 1/2$ with ε small enough to satisfy both conditions, and find a positive integer $M_\varepsilon \leq M_{\sigma^2}$ and constants $p_{1,\varepsilon}$ and $p_{2,\varepsilon}$ such that the pdf $f_{\varepsilon,N,\delta}$ defined by

$$\begin{aligned} f_{\varepsilon,N,\delta} =\,& p_{1,\varepsilon}[\Pi_\varepsilon(x+M_\varepsilon) + \Pi_\varepsilon(x+M_\varepsilon-2) \\ &+ \cdots + \Pi_\varepsilon(x-M_\varepsilon)] + p_{2,\varepsilon}[\Pi_\varepsilon(x+M_\varepsilon-1) \\ &+ \Pi_\varepsilon(x+M_\varepsilon-3) + \cdots + \Pi_\varepsilon(x-M_\varepsilon+1)] \end{aligned}$$

with

$$(M_\varepsilon+1)p_{1,\varepsilon} + M_\varepsilon p_{2,\varepsilon} = 1$$

exhibits second moment σ^2. Since ε was chosen small enough so that the second moment of $\Pi_\varepsilon(x)$ is less than σ^2, and since the second moment corresponding to the choice $M_\varepsilon = M_{\sigma^2}, p_{1,\varepsilon} = p_1, p_{2,\varepsilon} = p_2$ (where p_1 and p_2 correspond with $\mathcal{P}^{SV}_{\sigma^2}$) is strictly greater than σ^2, we remark that there exists such a $f_{\varepsilon,N,\delta}$ by noting that we can vary the second moment continuously between these two extremes by varying $M_\varepsilon, p_{1,\varepsilon}$ and $p_{2,\varepsilon}$.

Having constructed the pdf $f_{\varepsilon,N,\delta}$ to meet the power constraint (5), we note that for any $\varepsilon > 0$ the divergence $D(f_{\varepsilon,N,\delta}\|f_N)$ is finite, rendering $f_{\varepsilon,N,\delta}$ a valid candidate for (3) for sufficiently large δ.

Finally, we define a probability mass function $\mathcal{P}_\varepsilon$ corresponding to $f_{\varepsilon,N,\delta}$ by setting $\mathcal{P}_\varepsilon$ to take the value $p_{1,\varepsilon}$ on the points $-M_\varepsilon, -M_\varepsilon+2, \cdots, M_\varepsilon$ and the value $p_{2,\varepsilon}$ on the points

$-M_\varepsilon+1, -M_\varepsilon+3, \cdots, M_\varepsilon-1$ and note that the difference in second moments is given by

$$\int_{-M_\varepsilon-\varepsilon}^{M_\varepsilon+\varepsilon} x^2 f_{\varepsilon,N,\delta}(x)\,dx - \sum_{i=-M_\varepsilon}^{M_\varepsilon} i^2 \mathcal{P}_\varepsilon(i) = \frac{\varepsilon^2}{3}.$$

Hence, by taking ε small enough we can force the second moment of $f_{\varepsilon,N,\delta}$ arbitrarily close to that of $\mathcal{P}_\varepsilon$. By the fact that σ^2 uniquely determines M_{σ^2}, p_1 and p_2 for $\mathcal{P}_{\sigma^2}^{SV}$, we have that $M_\varepsilon \to M_{\sigma^2}, p_{1,\varepsilon} \to p_1$ and $p_{2,\varepsilon} \to p_2$ as $\varepsilon \to 0$, implying that $F_{\varepsilon,N,\delta}$ converges to $F_{\sigma^2}^{SV}$ as $\varepsilon \to 0$, and hence that the ML detection error probability $P_{\mathrm{ML}}(f_{\varepsilon,N,\delta})$ converges to that exhibited by $\mathcal{P}_{\sigma^2}^{SV}$. Since worst case performance $P_{\mathrm{ML}}(f_{\varepsilon,N,\delta})$ is lower-bounded by that of the explicit construction $f_{\varepsilon,N,\delta}$ and is upper-bounded by the performance exhibited by the unique solution $\mathcal{P}_{\sigma^2}^{SV}$ to the power-constrained problem, we conclude that

$$\lim_{\delta\to\infty} F_{\varepsilon,N,\delta} = F_{\sigma^2}^{SV},$$

the desired result. $\square$

D. Proof of Lemma 3

Choose any suitable nominal pdf f_N. The family $\mathcal{M}(a)$ of worst case amplitude-constrained noise pdf's clearly provides an upper bound for (3) since the amplitude-constrained problem corresponds with (3) after dropping the power constraint and replacing the divergence constraint with the weaker absolute continuity constraint for nominals with finite support $(-a, a)$. Note that the class of pulse-series candidate pdf's parameterized by $\varepsilon > 0$ and given by

$$f_\varepsilon(x) = \frac{1}{\lfloor a \rfloor + 1} \sum_{i=0}^{\lfloor a \rfloor} \Pi_\varepsilon(x + \lfloor a \rfloor - 2i)$$

falls into the class $\mathcal{M}(a)$ for $\varepsilon < a - \lfloor a \rfloor$, and that f_ε exhibits finite divergence with respect to f_N whenever $\varepsilon > 0$. Furthermore, the second moment of f_ε can be forced arbitrarily close to that of the probability mass function $\mathcal{P}_{\lfloor a \rfloor}$ taking the value $1/(\lfloor a \rfloor + 1)$ at the points $-\lfloor a \rfloor, -\lfloor a \rfloor + 2, \cdots, \lfloor a \rfloor$ by taking ε arbitrarily small. Since $a \le M_{\sigma^2}$, where M_{σ^2} represents the largest point taking mass in the mixture $\mathcal{P}_{\sigma^2}^{SV}$ of equiprobable distributions discussed above, and since $\lfloor a \rfloor \le a$, the second moment of $\mathcal{P}_{\lfloor a \rfloor}$ is strictly smaller than σ^2 by an analysis of such equiprobable mixtures provided in [13].

Hence, there exists $\varepsilon > 0$ small enough so that f_ε is a feasible solution to (3) for sufficiently large δ. By defining

$$\delta_{N,\sigma^2} = \min_{f_\mathcal{N} \in \mathcal{M}(a):\ \int_{-a}^{a} x^2 f_\mathcal{N}(x)\,dx \le \sigma^2} D(f_\mathcal{N} \| f_N)$$

and noting that all worst case amplitude-constrained pdfs belong to the class $\mathcal{M}(a)$, we have the result. $\square$

E. Proof of Lemma 4

Assume first that $a > \lceil \sqrt{3\sigma^2+1} - 1 \rceil$. Fix the divergence tolerance $\delta = \delta_0$ in (3) and assume by contradiction that there exists an optimal (worst case) solution $f_{\hat{N},N,\delta_0}$ with

$$D(f_{\hat{N},N,\delta_0} \| f_N) = \delta_1 < \delta_0.$$

Since it was shown in [13] that $\mathcal{P}_{\sigma^2}^{SV}$ is the unique solution to the optimization problem obtained by dropping the divergence constraint from (3), we know that $P_{\mathrm{ML}}(f_{\hat{N},N,\delta_0})$ falls strictly short of the ML detection error probability achieved by $\mathcal{P}_{\sigma^2}^{SV}$. We also know from Theorem 1 that $f_{\hat{N},N,\delta}$ tends to $\mathcal{P}_{\sigma^2}^{SV}$ in distribution as $\delta \to \infty$, implying that $P_{\mathrm{ML}}(f_{\hat{N},N,\delta})$ tends to the ML detection error probability achieved by $\mathcal{P}_{\sigma^2}^{SV}$. Hence, there exists some $\delta_2 > \delta_1$ such that

$$P_{\mathrm{ML}}(f_{\hat{N},N,\delta_2}) > P_{\mathrm{ML}}(f_{\hat{N},N,\delta_0}).$$

Define the candidate pdf

$$f'_{\hat{N},N,\delta_0} = \alpha f_{\hat{N},N,\delta_2} + (1-\alpha) f_{\hat{N},N,\delta_0}$$

where $0 < \alpha < 1$ and α is chosen small enough so that $D(f'_{\hat{N},N,\delta_0} \| f_N) < \delta_0$, which is possible as a result of the convexity of the divergence measure and the assumption $\delta_1 < \delta_0$. By convexity of the feasible set and concavity of the objective function associated with (3), $f'_{\hat{N},N,\delta_0}$ is necessarily a feasible solution which strictly outperforms $f_{\hat{N},N,\delta_0}$, rendering the latter suboptimal and thereby leading to the desired contradiction.

For the case $a \le \lceil \sqrt{3\sigma^2+1} - 1 \rceil$, fix $\delta_0 \le \delta_{N,\sigma^2}$ and assume by contradiction that there exists an optimal solution $f_{\hat{N},N,\delta_0}$ to (3) with

$$D(f_{\hat{N},N,\delta_0} \| f_N) = \delta_1 < \delta_0.$$

Since $\mathcal{M}(a)$ represents the class of maximizing solutions to the amplitude-constrained problem which in turn provides an upper bound for (3), and since it was shown in Lemma 3 that $\delta_{N,\sigma^2} < \infty$, there exists a divergence tolerance $\delta_2 > \delta_1$ and a corresponding solution $f_{\hat{N},N,\delta_2}$ for which

$$P_{\mathrm{ML}}(f_{\hat{N},N,\delta_2}) > P_{\mathrm{ML}}(f_{\hat{N},N,\delta_0}).$$

As above, the candidate pdf

$$f'_{\hat{N},N,\delta_0} = \alpha f_{\hat{N},N,\delta_2} + (1-\alpha) f_{\hat{N},N,\delta_0}$$

with $0 < \alpha < 1$ and α chosen small enough so that

$$D(f'_{\hat{N},N,\delta_0} \| f_N) \le \delta_0$$

is a feasible solution to (3) which strictly outperforms $f_{\hat{N},N,\delta_0}$, leading to the desired contradiction. $\square$

F. Proof of Lemma 5

Consider any solution $f_{\hat{N}}$ to (3). In the event that

$$D(f_{\hat{N}} \| f_N) = \delta_1 < \delta$$

we can construct a family of pdf's $f'_{\hat{N},t}$ parameterized by t for which the limiting distribution $F'_{\hat{N}} = \lim_{t\to\infty} F'_{\hat{N},t}$ exhibits divergence δ with respect to f_N and is also a solution as follows.

Let

$$\hat{x}_+ = \inf\{x > 0 : f_N(x) > 0\}$$
$$\hat{x}_- = \sup\{x < 0 : f_N(x) > 0\}$$

and

$$\hat{x} = \arg\min |x|, \qquad x \in \{\hat{x}_+, \hat{x}_-\}.$$

Intuitively, $\hat{x}$ represents a point to which we can arbitrarily transport mass within $f_{\hat{N}}$ without increasing its second moment, since the support of $f_{\hat{N}}$ is restricted to that of f_N and $\hat{x}$ is the point of minimum amplitude among the closure of such support. Assume without loss of generality that $\hat{x} \geq 0$ and consider the family of pdf's given by

$$f'_{\hat{N},t}(x) = \begin{cases} e^{\delta_0 t}, & 0 < x - \hat{x} < \eta_t \\ (1-k_t) f_{\hat{N}}(x), & \text{otherwise} \end{cases}$$

where $\delta_0 > 0, \eta_t = e^{-\delta_0 t}/t, t$ is chosen large enough so that $\hat{x} + \eta_t$ falls within the support of f_N, and k_t is determined by the unit integral of $f'_{\hat{N},t}$ according to

$$k_t = \frac{\frac{1}{t} - \int_{\hat{x}}^{\hat{x}+\eta_t} f_{\hat{N}}(x)\,dx}{1 - \int_{\hat{x}}^{\hat{x}+\eta_t} f_{\hat{N}}(x)\,dx}.$$

Then, for large enough t, $f'_{\hat{N},t}$ has second moment bounded above by that of $f_{\hat{N}}$ by the choice of $\hat{x}$, and exhibits divergence

$$\begin{aligned} D(f'_{\hat{N},t} \| f_N) &= (1-k_t)\log(1-k_t) + (1-k_t) D(f_{\hat{N}} \| f_N) \\ &\quad - \int_{\hat{x}}^{\hat{x}+\eta_t} (1-k_t) f_{\hat{N}}(x) \\ &\quad \cdot \log \frac{(1-k_t) f_{\hat{N}}(x)}{f_N(x)}\,dx \\ &\quad + \int_{\hat{x}}^{\hat{x}+\eta_t} e^{\delta_0 t} \log \frac{e^{\delta_0 t}}{f_N(x)}\,dx \\ &= (1-k_t)\log(1-k_t) + (1-k_t)\delta_1 + \delta_0 \\ &\quad - \int_{\hat{x}}^{\hat{x}+\eta_t} (1-k_t) f_{\hat{N}}(x) \\ &\quad \cdot \log \frac{(1-k_t) f_{\hat{N}}(x)}{f_N(x)}\,dx \\ &\quad - \int_{\hat{x}}^{\hat{x}+\eta_t} e^{\delta_0 t} \log f_N(x)\,dx. \end{aligned} \tag{39}$$

The last two terms in (39) both tend to zero as $t \to \infty$, as does the quantity k_t. Hence, the divergence of the limiting distribution $F'_{\hat{N}}$ with respect to the distribution F_N is given by

$$D(F'_{\hat{N}} \| F_N) = \int_{-\infty}^{\infty} \log \frac{dF'_{\hat{N}}}{dF_N}\,dF_N = \delta_1 + \delta_0$$

allowing satisfaction of the divergence constraint through the choice $\delta_0 = \delta - \delta_1$. Finally, a simple lower bound on the performance of $f'_{\hat{N},t}$ is given by

$$\begin{aligned} P_{\mathrm{ML}}(f'_{\hat{N},t}) &= \int_{-a}^{a} \min\{f'_{\hat{N},t}(x), f'_{\hat{N},t}(x+2)\}\,dx \\ &\geq \int_{\{x\in[-a,a]:\, x\notin[\hat{x}-2,\hat{x}+\eta_t-2]\cup[\hat{x},\hat{x}+\eta_t]\}} \\ &\quad \cdot \min\{f'_{\hat{N},t}(x), f'_{\hat{N},t}(x+2)\}\,dx \\ &= (1-k_t)\left(\int_{-a}^{a} \min\{f_{\hat{N}}(x), f_{\hat{N}}(x+2)\}\,dx\right. \\ &\quad - \int_{\hat{x}-2}^{\hat{x}+\eta_t-2} \min\{f_{\hat{N}}(x), f_{\hat{N}}(x+2)\}\,dx \\ &\quad \left. - \int_{\hat{x}}^{\hat{x}+\eta_t} \min\{f_{\hat{N}}(x), f_{\hat{N}}(x+2)\}\,dx\right) \\ &\geq (1-k_t)\left(\int_{\hat{x}-2}^{\hat{x}} P_{\mathrm{ML}}(f_{\hat{N}})\right. \\ &\quad \left. - 2\int_{\hat{x}}^{\hat{x}+\eta_t} f_{\hat{N}}(x)\,dx\right) \end{aligned}$$

which converges to $P_{\mathrm{ML}}(f_{\hat{N}})$ with growing t, so that performance of the limiting distribution $F'_{\hat{N}}$ matches that of F_N, rendering the former a valid worst case solution. $\square$

G. Proof of Lemma 6

We prove that given a symmetric continuous nominal f_N with support $(-a, a)$ and a divergence tolerance δ for which there exists $\delta_2 > \delta$ such that $P_{\mathrm{ML}}(f_{\hat{N},N,\delta_2}) > P_{\mathrm{ML}}(f_{\varepsilon,N,\delta})$, the worst case pdf $f_{\varepsilon,N,\delta}$ is continuous over the interval $(-a, a)$. Take the notation-simplifying convention that $f_{\hat{N}}$ represents the worst case pdf $f_{\hat{N},N,\delta}$ and assume by contradiction that $f_{\hat{N}}$ exhibits at least one discontinuity at some point $x_0 \in (-a, a)$. We assume that the discontinuities of $f_{\hat{N}}$ are isolated, which is justified by the observation that for any pdf with nonisolated discontinuities, there exists another in the equivalence class determined by equality up to measure zero with isolated discontinuities and sharing equal values for the objective function and constraints. Assume without loss of generality that $f_{\hat{N}}(x_0^-) < f_{\hat{N}}(x_0^+)$, where, in general, $f(x_0^-)$ and $f(x_0^+)$ represent

$$\lim_{x\to x_0, x < x_0} f(x)$$

and

$$\lim_{x\to x_0, x > x_0} f(x)$$

respectively, definitions justified by the assumption of isolated discontinuities.

Define

$$\begin{aligned} \underline{M} = \min\{ & i = 0,1,2,\cdots: f_{\hat{N}}(x_0 - 2i^+) \\ & > f_{\hat{N}}(x_0 - 2i - 2^+) \\ & \text{or } f_{\hat{N}}(x_0 - 2i^-) < f_{\hat{N}}(x_0 - 2i - 2^-)\} \\ \overline{M} = \min\{ & i = 0,1,2,\cdots: f_{\hat{N}}(x_0 + 2i^+) \\ & > f_{\hat{N}}(x_0 + 2i + 2^+) \\ & \text{or } f_{\hat{N}}(x_0 + 2i^-) < f_{\hat{N}}(x_0 + 2i + 2^-)\}. \end{aligned}$$

The finiteness of $\underline{M}$ and $\overline{M}$ are guaranteed by the integrability of $f_{\hat{N}}$. Note that there is necessarily a discontinuity at each of the points $x_0 + 2i, i = -\underline{M}, -\underline{M}+1, \cdots, \overline{M}$ with magnitude at least as great as the discontinuity at x_0.

Consider the class of candidate pdf's parameterized by $\xi > 0,\ \varepsilon > 0$ and given by

$$f_{\xi,\varepsilon}(x) = \begin{cases} f_{\hat{N}}(x) + \varepsilon, & x_0 + 2i - \xi < x < x_0 + 2i, \\ & i = -\underline{M}, -\underline{M}+1, \cdots, \overline{M} \\ f_{\hat{N}}(x) - \varepsilon, & x_0 + 2i \leq x < x_0 + 2i + \xi, \\ & i = -\underline{M}, -\underline{M}+1, \cdots, \overline{M} \\ f_{\hat{N}}(x), & \text{otherwise} \end{cases}$$

where ξ is chosen small enough so that the discontinuities of $f_{\hat{N}}$ within the intervals $[x_0+2i-\xi, x_0+2i+\xi]$ are restricted to the points $x_0+2i, i=-\underline{M}, -\underline{M}+1, \cdots, \overline{M}$ (justified by the assumption that $f_{\hat{N}}$ has isolated discontinuities), and ξ and ε are chosen small enough so that either $f_{\xi,\varepsilon}(x) < f_{\xi,\varepsilon}(x-2)$ throughout $[x_0-2\underline{M}-\xi, x_0-2\underline{M})$ or $f_{\xi,\varepsilon}(x) > f_{\xi,\varepsilon}(x-2)$ throughout $[x_0-2\underline{M}, x_0-2\underline{M}+\xi]$ and so that either $f_{\xi,\varepsilon}(x) < f_{\xi,\varepsilon}(x+2)$ throughout $[x_0+2\overline{M}-\xi, x_0+2\overline{M})$ or $f_{\xi,\varepsilon}(x) > f_{\xi,\varepsilon}(x+2)$ throughout $[x_0+2\overline{M}, x_0+2\overline{M}+\xi]$. Such a specification is always possible given the definitions of $\underline{M}$ and $\overline{M}$, and ensures that $P_{\mathrm{ML}}(f_{\xi,\varepsilon}) \geq P_{\mathrm{ML}}(f_{\hat{N}})$. Furthermore, restrict ξ and ε small enough so that for each

$$i=-\underline{M}, -\underline{M}+1, \cdots, \overline{M} \qquad f_{\xi,\varepsilon}(x) < f_{\xi,\varepsilon}(y)$$

for all x in $[x_0+2i-\xi, x_0+2i)$ and y in $[x_0+2i, x_0+2i+\xi]$. Finally, fix $\gamma > 0$ and require that ξ and ε are small enough so that

$$\min_{i=-\underline{M}, -\underline{M}+1, \cdots, \overline{M}} \left\{ \inf_{x\in[x_0+2i-\xi, x_0+2i), y\in[x_0+2i, x_0+2i+\xi)} \cdot \left\{ \log \frac{f_{\xi,\varepsilon}(y)}{f_N(y)} - \log \frac{f_{\xi,\varepsilon}(x)}{f_N(x)} \right\} \right\} \geq \gamma$$

which is always achievable for small enough γ given that f_N is assumed continuous, particularly at the points x_0+2i, $i=-\underline{M}, -\underline{M}+1, \cdots, \overline{M}$. The displacement of mass around the point x_0 results in a reduction in divergence given by

$$\begin{aligned} &\int_{x_0-\xi}^{x_0} \left[(f_{\hat{N}}(x)+\varepsilon) \log \frac{f_{\hat{N}}(x)+\varepsilon}{f_N(x)} - f_{\hat{N}}(x) \log \frac{f_{\hat{N}}(x)}{f_N(x)} \right] dx \\ &\quad + \int_{x_0}^{x_0+\xi} \left[(f_{\hat{N}}(x)-\varepsilon) \log \frac{f_{\hat{N}}(x)-\varepsilon}{f_N(x)} - f_{\hat{N}}(x) \log \frac{f_{\hat{N}}(x)}{f_N(x)} \right] dx \\ &= \int_{x_0-\xi}^{x_0} \left[f_{\hat{N}}(x) \log \left(1+\frac{\varepsilon}{f_{\hat{N}}(x)}\right) + \varepsilon \log \frac{f_{\hat{N}}(x)+\varepsilon}{f_N(x)} \right] dx \\ &\quad + \int_{x_0}^{x_0+\xi} \left[f_{\hat{N}}(x) \log \left(1-\frac{\varepsilon}{f_{\hat{N}}(x)}\right) - \varepsilon \log \frac{f_{\hat{N}}(x)-\varepsilon}{f_N(x)} \right] dx \\ &\leq \varepsilon \left[\int_{x_0-\xi}^{x_0} \log \frac{f_{\hat{N}}(x)+\varepsilon}{f_N(x)}\, dx - \int_{x_0}^{x_0+\xi} \log \frac{f_{\hat{N}}(x)-\varepsilon}{f_N(x)}\, dx \right] \\ &\leq -\gamma\varepsilon\xi \end{aligned} \tag{40}$$

where the convention is used that $f_{\hat{N}}(x)\log(1+\varepsilon/f_{\hat{N}}(x))=0$ whenever $f_{\hat{N}}(x)=0$ and where (40) follows from the specification that $f_{\xi,\varepsilon}(y) > f_{\xi,\varepsilon}(x)$ for all x in $[x_0-\xi, x_0]$ and y in $[x_0, x_0+\xi]$, and from the construction of $f_{\xi,\varepsilon}$ from $f_{\hat{N}}$. Similarly, the displacement of mass around each of the points $x_0+2i, i\neq 0$ results in a reduction in divergence, with the overall result being that $D(f_{\xi,\varepsilon}\|f_N) - D(f_{\hat{N}}\|f_N) \leq -\gamma\varepsilon\xi$ where $\gamma > 0$ is fixed.

The construction $f_{\xi,\varepsilon}$ has been shown to match or outperform $f_{\hat{N}}$ in the objective function while exhibiting a strictly smaller divergence with respect to the nominal f_N. In light of the assumed activity of the divergence constraint, we could immediately claim suboptimality of $f_{\hat{N}}$ if it were not for the possible increase in second moment incurred by the construction of $f_{\xi,\varepsilon}$ from $f_{\hat{N}}$, which we deal with presently. An upper bound on such increase is obtained by assuming the worst case scenario that the displacement of mass around each of the points $x_0+2i, i=-\underline{M}, -\underline{M}+1, \cdots, \overline{M}$ leads to an increased second moment and that each point takes the value of the upper bound $|x_0|+\underline{M}+\overline{M}$ on magnitude, and is given by

$$\begin{aligned} &\int_{-a}^{a} x^2 (f_{\xi,\varepsilon}(x) - f_{\hat{N}}(x))\, dx \\ &\quad \leq (\underline{M}+\overline{M}+1) \left[\int_{|x_0|+\underline{M}+\overline{M}}^{|x_0|+\underline{M}+\overline{M}+\xi} \varepsilon x^2\, dx - \int_{|x_0|+\underline{M}+\overline{M}-\xi}^{|x_0|+\underline{M}+\overline{M}} \varepsilon x^2 dx \right] \\ &\quad = 2(\underline{M}+\overline{M}+1)(|x_0|+\underline{M}+\overline{M})\varepsilon\xi^2. \end{aligned} \tag{41}$$

In order to satisfy the power constraint, we take advantage of the fact that the reduction in divergence is $O(\varepsilon\xi)$ while the increase in second moment is $O(\varepsilon\xi^2)$. Take any $\delta_2 > \delta$ for which $P_{\mathrm{ML}}(f_{\hat{N},N,\delta_2}) > P_{\mathrm{ML}}(f_{\hat{N}})$, a specification supported by the original assumptions. Find any two disjoint intervals $[x_1, x_1+v]$ and $[x_2, x_2+v], 0 \leq x_1 < x_2, 0 < v < x_2 - x_1$ of positive support for $f_{\hat{N},N,\delta_2}$. Now construct the pdf

$$f_{v,w}(x) = \begin{cases} f_{\hat{N},N,\delta_2}(x)+w, & x\in[x_1, x_1+v] \\ f_{\hat{N},N,\delta_2}(x)-w, & x\in[x_2, x_2+v] \\ f_{\hat{N},N,\delta_2}(x), & \text{otherwise} \end{cases}$$

where $w > 0$ and v, w are chosen small enough so that $P_{\mathrm{ML}}(f_{v,w}) > P_{\mathrm{ML}}(f_{\hat{N}})$, which is always possible since $P_{\mathrm{ML}}(f_{\hat{N},N,\delta_2}) > P_{\mathrm{ML}}(f_{\hat{N}})$. Note that

$$\begin{aligned} x^2 f_{v,w}(x)\, dx &= \sigma^2 - vw[(x_2^2 - x_1^2) + w(x_2 - x_1)] \\ &\leq \sigma^2 - vw(x_2^2 - x_1^2). \end{aligned} \tag{42}$$

Finally, construct the new candidate pdf

$$f'_{\hat{N}}(x) = (1-\varepsilon\xi^{\frac{3}{2}} f_{\xi,\varepsilon}(x) + \varepsilon\xi^{\frac{3}{2}} f_{v,w}(x)$$

noting that

$$\begin{aligned} P_{\mathrm{ML}}(f'_{\hat{N}}) &\geq (1-\varepsilon\xi^{\frac{3}{2}} P_{\mathrm{ML}}(f_{\xi,\varepsilon}) + \varepsilon\xi^{\frac{3}{2}} P_{\mathrm{ML}}(f_{v,w}) \\ &> P_{\mathrm{ML}}(f_{\hat{N}}). \end{aligned}$$

Define $\delta_3 = D(f_{v,w}\|f_N)$ and note that

$$\begin{aligned} D(f'_{\hat{N}}\|f_N) &< (1-\varepsilon\xi^{\frac{3}{2}})(\delta - \gamma\varepsilon\xi) + \varepsilon\xi^{\frac{3}{2}}\delta_3 \\ &= \delta - \gamma\varepsilon\xi + o(\varepsilon\xi) \end{aligned}$$

by the convexity of the divergence measure, implying that the divergence constraint is satisfied for small enough values of ε

and ξ. The second moment of $f'_{\hat{N}}p$ satisfies

$$\begin{aligned}\int_{-a}^{a} x^2 f'_{\hat{N}}(x)\,dx &= (1-\varepsilon\xi^{\frac{3}{2}})\int_{-a}^{a} x^2 f_{\xi,\varepsilon}(x)\,dx \\ &\quad + \varepsilon\xi^{\frac{3}{2}}\int_{-a}^{a} x^2 f_{v,w}(x)\,dx \\ &< (1-\varepsilon\xi^{\frac{3}{2}})(\sigma^2+2(\underline{M}+\overline{M}+1) \\ &\quad \cdot(|x_0|+\underline{M}+\overline{M})\varepsilon\xi^2) \\ &\quad +\varepsilon\xi^{\frac{3}{2}}(\sigma^2-vw(x_2^2-x_1^2)) \\ &= \sigma^2-vw(x_2^2-x_1^2)\varepsilon\xi^{\frac{3}{2}}+o(\varepsilon\xi^{\frac{3}{2}}) \quad (43)\end{aligned}$$

where (43) follows from (41) and (42), implying further that $f'_{\hat{N}}p$ satisfies the power constraint for small enough ε and ξ. Hence, the construction of $f'_{\hat{N}}$ demonstrates suboptimality of the discontinuous candidate $f_{\hat{N}}$, providing the desired contradiction. The case $f_{\hat{N}}(x_0^-) > f_{\hat{N}}(x_0^+)$ can be treated with a trivial modification of the above argument. □

H. Proof of Lemma 7

The zero-threshold detector (hard limiter) yields the decision H_0 if $Y<0$ and H_1 if $Y\geq 0$ with reference to the hypothesis test (1). This decision strategy falls into the class of ML strategies if and only if the symmetric worst case pdf $f_{\hat{N}}$ exhibits $f_{\hat{N}}(x)\geq f_{\hat{N}}(x+2)$ for all $x\geq -1$, except perhaps on a set of measure zero. Assume by contradiction, then, that the worst case pdf $f_{\hat{N}}$ exhibits $f_{\hat{N}}(x)<f_{\hat{N}}(x+2)$ over some set of positive measure in $[-1,\infty)$. Such a set must have finite measure as a result of the integrability of $f_{\hat{N}}$, and can hence be approximated arbitrarily well by a finite union of open intervals (as discussed in [24, p. 72] and elsewhere). Pick any such interval $(x^*, x^*+\eta)$ where $x^*\geq -1$ and $\eta>0$, and define the candidate pdf

$$f'_{\hat{N},\varepsilon}(x)=\begin{cases} f_{\hat{N}}(x)+\varepsilon, & x\in(x^*,x^*+\eta)\\ f_{\hat{N}}(x)-\varepsilon, & x\in(x^*+2,x^*+2+\eta)\\ f_{\hat{N}}(x), & \text{otherwise}\end{cases}$$

where $\varepsilon>0$ is chosen small enough so that

$$f'_{\hat{N},\varepsilon}(x)<f'_{\hat{N},\varepsilon}(x+2)$$

over $(x^*,x^*+\eta)$. The construction $f'_{\hat{N},\varepsilon}$ exhibits second moment strictly less than that of $f_{\hat{N}}$, and it is easy to verify that the ML criterion (2) satisfies $P_{\text{ML}}(f'_{\hat{N},\varepsilon})\geq P_{\text{ML}}(f_{\hat{N}})$ given the restriction on ε. A comparison of divergences yields

$$\begin{aligned}&D(f'_{\hat{N},\varepsilon}\|f_N)-D(f_{\hat{N}}\|f_N)\\ &=\int_{x^*}^{x^*+\eta}\Bigg[(f_{\hat{N}}(x)+\varepsilon)\log\frac{f_{\hat{N}}(x)+\varepsilon}{f_N(x)}-f_{\hat{N}}(x)\\ &\quad\cdot\log\frac{f_{\hat{N}}(x)}{f_N(x)}+(f_{\hat{N}}(x+2)-\varepsilon)\log\frac{f_{\hat{N}}(x+2)-\varepsilon}{f_N(x+2)}\\ &\quad - f_{\hat{N}}(x+2)\log\frac{f_{\hat{N}}(x+2)-\varepsilon}{f_N(x+2)}\\ &\quad - f_{\hat{N}}(x+2)\frac{f_{\hat{N}}(x+2)}{f_N(x+2)}\Bigg]\,dx\\ &=\int_{x^*}^{x^*+\eta}\Bigg[f_{\hat{N}}(x)\log\left(1+\frac{\varepsilon}{f_{\hat{N}}(x)}\right)\\ &\quad+\varepsilon\log\frac{f_{\hat{N}}(x)+\varepsilon}{f_N(x)}+f_{\hat{N}}(x+2)\\ &\quad\cdot\log\left(1-\frac{\varepsilon}{f_{\hat{N}}(x+2)}\right)-\varepsilon\log\frac{f_{\hat{N}}(x+2)-\varepsilon}{f_N(x+2)}\Bigg]\\ &\quad\cdot dx \qquad (44)\\ &\leq\varepsilon\int_{x^*}^{x^*+\eta}\log\frac{(f_{\hat{N}}(x)+\varepsilon)f_N(x+2)}{(f_{\hat{N}}(x+2)-\varepsilon)f_N(x)}\,dx \qquad (45)\\ &<0 \qquad (46)\end{aligned}$$

where the convention is used in (44) that

$$f_{\hat{N}}(x)\log(1+\varepsilon/f_{\hat{N}}(x))=0$$

when $f_{\hat{N}}(x)=0$, where (45) follows from the relationship $\log x\leq x-1$, and where (46) follows from the assumption that f_N is unimodal and from the previous restriction on ε. Hence, for small enough ε the construction $f'_{\hat{N},\varepsilon}$ strictly outperforms $f_{\hat{N}}$ in the power and divergence constraints, while matching or outperforming $f_{\hat{N}}$ in the objective function. In the event $a>\lceil\sqrt{3\sigma^2+1}-1\rceil$, the divergence constraint is active as a result of Lemma 4. In the event $a\leq\lceil\sqrt{3\sigma^2+1}-1\rceil$, the divergence constraint is also active in view of Lemma 4 and the observation that the worst case pdf $f_{\hat{N}}$ is assumed not to be 2-periodic, which is a requirement for all pdf's in $\mathcal{M}(a)$. Hence, the strict divergence discrepancy between $f'_{\hat{N},\varepsilon}$ and $f_{\hat{N}}$ renders the latter suboptimal, providing the desired contradiction. □

I. Proof of Lemma 8

It is straightforward to show that, as a direct result of condition (13) and the assumed symmetry of f_N, the variance-scaled nominal $f_{\tilde{N}}$ is strictly unimodal (i.e., $f_{\tilde{N}}(x_1)>f_{\tilde{N}}(x_2)$ for all $0\leq x_1<x_2\leq a$). The following observation is key to the result at hand.

Lemma I.1: Let

$$I_n=[n-1+\xi_1,n-1+\xi_2],\qquad 0\leq\xi_1<\xi_2<1$$

represent any interval of increase for $f_{\hat{N}}$ contained in an interval $(n-1,n]$ where $n\geq 1$ is any positive integer, and where such increase is not necessarily strict. Assume furthermore that the interval

$$I_{n+2}=[n+1+\xi_1,n+1+\xi_2]$$

is an interval of strict decrease for $f_{\hat{N}}$. Then I_n is contained in $\mathcal{C}_m,\overline{\mathcal{C}}_m,\mathcal{C}_m^L$, or $\mathcal{C}_m^R$ where $m>n$ in the optimal form (11).

Proof: With reference to (11), if the interval I_n is contained in $\mathcal{A},\mathcal{B}$, or $\overline{\mathcal{B}}$, then $f_{\hat{N}}$ is a scaled version of the strictly unimodal pdf $f_{\tilde{N}}$ over I_n and is hence strictly decreasing within I_n. Furthermore, I_n is a subset of one of the sets of Λ since $f_{\hat{N}}$ is nonzero over I_n and $f_{\hat{N}}$ does not agree over I_n and I_{n+2}. Thus I_n belongs to $\mathcal{C}_m,\overline{\mathcal{C}}_m,\mathcal{C}_m^L$, or $\mathcal{C}_m^R$ for some $m\geq 2$. Now assume by contradiction that $m\leq n$. According to the optimal form (11), $f_{\hat{N}}$ takes the form

$$f_{\hat{N}}(x)=K\left(\prod_{i=0}^{m-1}f_{\tilde{N}}(x-2i)\right)^{\frac{1}{m}}$$

over I_n, where K is a scaling constant. If $m \le \lceil n/2 \rceil$, then

$$\begin{aligned} x - 2(m-1) &\ge x - 2\left\lceil \frac{n}{2} \right\rceil + 2 \\ &\ge x - n + 1 \\ &\ge 0 \end{aligned}$$

since x falls in $(n-1, n]$, and $f_{\hat{N}}(x)$ is therefore strictly decreasing over I_n as a result of the strict unimodality of $f_{\bar{N}}$. If, on the other hand, $\lceil \frac{n}{2} \rceil < m \le n$, we can rewrite

$$\begin{aligned} f_{\hat{N}}^m(x) &= K^m \prod_{i=0}^{2\lceil \frac{n}{2} \rceil - m - 1} f_{\bar{N}}(x-2i) \prod_{i=2\lceil \frac{n}{2} \rceil - m}^{m-1} f_{\bar{N}}(x-2i) \\ &= K^m \prod_{i=0}^{2\lceil \frac{n}{2} \rceil - m - 1} f_{\bar{N}}(x-2i) \prod_{i=\lceil \frac{n}{2} \rceil}^{m-1} f_{\bar{N}}(x-2i) \\ &\quad \times f_{\bar{N}}\left(x - 4\left\lceil \frac{n}{2} \right\rceil + 2 + 2i\right) \\ &= K^m \prod_{i=0}^{2\lceil \frac{n}{2} \rceil - m - 1} f_{\bar{N}}(x-2i) \prod_{i=\lceil \frac{n}{2} \rceil}^{m-1} f_{\bar{N}}(x-2i) \\ &\quad \times f_{\bar{N}}\left(4\left\lceil \frac{n}{2} \right\rceil - 2 - x - 2i\right) \qquad (47) \\ &= K^m \prod_{i=0}^{2\lceil \frac{n}{2} \rceil - m - 1} f_{\bar{N}}(x-2i) \prod_{i=\lceil \frac{n}{2} \rceil}^{m-1} \\ &\quad \times f_{\bar{N}}\left(2\left\lceil \frac{n}{2} \right\rceil - 1 - 2i + \left[x - 2\left\lceil \frac{n}{2} \right\rceil - 1\right]\right) \\ &\quad \times f_{\bar{N}}\left(2\left\lceil \frac{n}{2} \right\rceil - 1 - 2i - \left[x - 2\left\lceil \frac{n}{2} \right\rceil - 1\right]\right) \qquad (48) \end{aligned}$$

where (47) follows from the symmetry of $f_{\hat{N}}$ through Lemma 2. Note that $|x - 2\lceil \frac{n}{2} \rceil - 1| < 1$ since x falls in the interval $(n-1, n]$, implying that the right-most product in (48) is a strictly decreasing function of x as a direct result of condition (13). The remaining product in (48) is also a strictly decreasing function of x since $f_{\bar{N}}$ is strictly unimodal, implying that $f_{\hat{N}}(x)$ is strictly decreasing over I_n, providing the desired contradiction. $\square$

A direct result of Lemma I.1 is that any interval $[y_1, y_2]$, $y_1 \ge 0$ of increase for $f_{\hat{N}}(x)$ satisfies

$$f_{\hat{N}}(x) = f_{\hat{N}}(x-2) = \cdots = f_{\hat{N}}(x - 2\lceil y_2 \rceil) \qquad (49)$$

for all x in $[y_1, y_2]$ according to the definitions of $\mathcal{C}_m, \overline{\mathcal{C}}_m, \mathcal{C}_m^L$, and $\mathcal{C}_m^R$. This in turn implies through symmetry that

$$f_{\hat{N}}(x) = f_{\hat{N}}(2\lceil y_2 \rceil - x) \qquad (50)$$

for all x in $[y_1, y_2]$; that is, *if* $f_{\hat{N}}(x)$ is increasing over an interval $[y_1, y_2], y_1 \ge 0$, then $f_{\hat{N}}(x)$ is symmetric about $\lceil y_2 \rceil$. This observation provides the key for parameterizing the support decomposition Λ.

Recalling the form (11) for $f_{\hat{N}}$, we note first that there must be at least one interval of increase (not necessarily strict) in the region $[0, a]$; that is, $f_{\hat{N}}$ is *not* strictly unimodal in spite of the strict unimodality of $f_{\bar{N}}$. The reasoning is that any support decomposition Λ which includes a set $\mathcal{C}_n, \overline{\mathcal{C}_n}, \mathcal{C}_n^L$, or $\mathcal{C}_n^R$ in the interval $[2, a]$ (assuming $a > 2$) requires that $f_{\hat{N}}(x) = f_{\hat{N}}(x-2)$ for some x in the same interval, necessitating an interval of increase as a result of the continuity of $f_{\hat{N}}$. If there exists a set $\mathcal{C}_n, \overline{\mathcal{C}_n}, \mathcal{C}_n^L$, or $\mathcal{C}_n^R$ in the interval $[0, 2]$, then

$$f_{\hat{N}}(x) = f_{\hat{N}}(x-2) = f_{\hat{N}}(2-x)$$

by symmetry of $f_{\hat{N}}$ for some x in $[0, 2]$, again necessitating an interval of increase since any measurable set $\mathcal{C}_n, \overline{\mathcal{C}}_n, \mathcal{C}_n^L$, or $\mathcal{C}_n^R$ spans more than the singleton $\{x = 1\}$. Hence, the decomposition Λ is restricted to the sets $\mathcal{A}, \mathcal{B}$, and $\overline{\mathcal{B}}$ if there is to be no interval of increase in $[0, a]$. Since the value of $f_{\hat{N}}(x)$ over each of these sets is a scalar weighting of $f_{\bar{N}}(x)$ according to (11), $f_{\hat{N}}$ must be identical to $f_{\bar{N}}$ by the continuity and symmetry of $f_{\hat{N}}$, implying that $f_{\hat{N}}$ is suboptimal by the assumed activity of the divergence constraint.

Now let x_1 be the supremum of all points of increase for $f_{\hat{N}}$, whose finiteness is guaranteed by the integrability of $f_{\hat{N}}$ through (49). Interestingly, x_1 must be an integer, since

$$f_{\hat{N}}(x_1) = f_{\hat{N}}(2\lceil x_1 \rceil - x_1)$$

according to (50), which by continuity directly contradicts the definition of x_1 unless $x_1 = \lceil x_1 \rceil$.

Now let x_2 be the supremum of those points less than x_1 of strict decrease for $f_{\hat{N}}$, and let x_3 in turn be the supremum of those points less than x_2 of increase for $f_{\hat{N}}$. Note that the interval $[x_2, x_1]$ is an interval of increase, while $[x_3, x_2]$ is an interval of strict decrease. Note further that $x_2 \ge x_1 - 1$ since any interval of increase in the interval $[x_1 - 2, x_1 - 1]$ exhibits symmetry about the point $x_1 - 1$ according to (50), which cannot be the case given that the entire interval $[x_1 - 1, x_1]$ is an interval of increase whenever $x_2 < x_1 - 1$. We, therefore, know from (50) that $f_{\hat{N}}(x_2) = f_{\hat{N}}(2x_1 - x_2)$, along with the fact that $f_{\hat{N}}(x_3) = f_{\hat{N}}(2\lceil x_3 \rceil - x_3)$. If $x_3 > x_1 - 1$ we have that $f_{\hat{N}}(2x_1 - x_2) < f_{\hat{N}}(2x_1 - x_3)$, again contradicting the defintion of x_1 since $2x_1 - x_3 > x_1$. Therefore, we have $x_3 \le x_1 - 1$, implying that the interval $[x_1 - 1, x_1]$ containts exactly one interval $[x_1 - 1, x_2]$ of strict decrease and exactly one interval $[x_2, x_1]$ of increase for $f_{\hat{N}}$. Furthermore, since $f_{\hat{N}}$ is symmetric about x_1 over the interval $[x_2, 2x_1 - x_2]$, we know from Lemma I.1 that $\mathcal{C}_{x_1+1} = [x_2, 2x_1 - x_2)$ in the optimal form (11).

Consider the case $x_1 = 1$, for which the above analysis yields that $\mathcal{C}_2 = [x_2, 2 - x_2]$ where $0 \le x_2 < 1$. Since there are no points of increase greater than $x = 1$, it follows directly that $\mathcal{A} = (-a, x_2 - 2) \cup [2 - x_2, a)$ and $\mathcal{B} = [-x_2, x_2)$ by the unimodality of $f_{\bar{N}}$ and the continuity of $f_{\hat{N}}$, proving the portion of Lemma 8 corresponding to the case $\lfloor \overline{x} \rfloor = 1$ upon substituting $\overline{x} = 2 - x_2$.

For the case $x_1 \ge 2$, consider the interval $[x_1 - 2, x_1 - 1]$. Recalling from (49) that

$$f_{\hat{N}}(x_2) = f_{\hat{N}}(x_2 - 2) = \cdots = f_{\hat{N}}(x_2 - 2x_1)$$

we have that

$$f_{\hat{N}}(x_2) = f_{\hat{N}}(x_2 - 2x_1 + 2) = f_{\hat{N}}(2x_1 - x_2 - 2)$$

by symmetry of $f_{\hat{N}}$; since $x_1 - 1 < x_2 \le x_1$, we have directly that $x_1 - 2 \le 2x_1 - x_2 - 2 < x_1 - 1$, and hence there must be at least one interval of increase in the interval $[x_1 - 2, x_1 - 1]$ by the continuity of $f_{\hat{N}}$. Let x_4 be the supremum of all points of increase for $f_{\hat{N}}(x)$ in this interval. Noting as before that $f_{\hat{N}}(x_4) = f_{\hat{N}}(2(x_1-1)-x_4)$ according to (50), we have necessarily that $x_4 = x_1 - 1$. Further, by the same reasoning we used for the interval $[x_1-1, x_1]$, there must be exactly one interval of strict decrease $[x_1-2, x_5]$, $x_1 - 2 < x_5 < x_1 - 1$ and exactly one interval of increase $[x_5, x_1-1]$ within the interval $[x_1-2, x_1-1]$, completely determining that $\mathcal{C}_{x_1} = [x_5, 2(x_1-1)-x_5]$. Symmetry of $f_{\hat{N}}$ within $\mathcal{C}_{x_1}$ about the point $x_1 - 1$ then determines that $x_2 = 2(x_1-1)-x_5$. Recalling the definition of $\mathcal{C}_n, f_{\hat{N}}(x)$ is completely determined up to a weighting factor over the intervals

$$[x_5 - 2i, x_2 - 2i), \qquad i = 0, 1, \cdots, x_1 - 1$$

and

$$[x_2 - 2i, 2x_1 - x_2 - 2i), \qquad i = 0, 1, \cdots, x_1$$

by the specifications of $\mathcal{C}x_1$ and $\mathcal{C}_{x_1+1}$, completely determining $f_{\hat{N}}(x)$ up to a weighting factor over the interval $[-(2x_1-x_2), 2x_1-x_2]$ by continuity. By the strict unimodality of $f_{\tilde{N}}$ and the continuity of $f_{\hat{N}}$, the remaining intervals

$$(-a, -(2x_1 - x_2)) \quad \text{and} \quad [2x_1 - x_2, a)$$

fall within the set $\mathcal{A}$, completing the proof of Lemma 8 upon substituting $\overline{x} = 2x_1 - x_2$. □

J. Proof of Lemma 9

Take any nominal f_N satisfying the conditions of Theorem 2. Assume by contradiction that the worst case noise pdf $f_{\hat{N}}$ exhibits second moment $\sigma_0^2 < \sigma^2$. Consider the worst case noise pdf $f_{\tilde{N}}$ yielded by a lone divergence constraint and no power constraint, a problem investigated more thoroughly in Section II-B, with the divergence tolerance set to δ. An expression (29) for $f_{\tilde{N}}$ is given by Theorem 3, which was developed without the need to consider a power constraint. The form of (29) reveals that the second moment of $f_{\tilde{N}}$ is strictly greater than σ^2 for any strictly positive divergence tolerance, implying by uniqueness that $P_{\rm ML}(f_{\tilde{N}}) > P_{\rm ML}(f_{\hat{N}})$ as expected. Now consider the new candidate pdf $f'_{\hat{N}}$ for (3) given by

$$f'_{\hat{N}}(x) = \alpha f_{\tilde{N}}(x) + (1-\alpha) f_{\hat{N}}(x)$$

where $0 < \alpha < 1$ and α is chosen small enough so that the second moment of $f'_{\hat{N}}$ is no greater than σ^2. By strict convexity of the divergence measure, $D(f'_{\hat{N}}||f_N) < \delta$, rendering $f'_{\hat{N}}$ a feasible solution for (3). Concavity of the objective function then implies that

$$P_{\rm ML}(f'_{\hat{N}}) \geq \alpha P_{\rm ML}(f_{\tilde{N}}) + (1-\alpha) P_{\rm ML}(f_{\hat{N}}) > P_{\rm ML}(f_{\hat{N}})$$

demonstrating sub-optimality of $f_{\hat{N}}$ and providing the desired contradiction. □

K. Proof of Lemma 15

Let the divergence tolerance take any value $\delta \leq \delta_N$ and assume by contradiction that the worst case pdf $f_{\tilde{N},N,\delta}$ exhibits

$$D(f_{\tilde{N},N,\delta}||f_N) = \delta_1 < \delta.$$

By definition of δ_N, $f_{\tilde{N},N,\delta}$ does not belong to the class $\mathcal{M}(a)$ and hence

$$P_{\rm ML}(f_{\tilde{N},N,\delta}) < \lfloor a \rfloor / (2(\lfloor a \rfloor + 1)).$$

Take any pdf $f_{\mathcal{N}}$ in $\mathcal{M}(a)$, which is a justifiable specification in view of the existence of a finite δ_{N,σ^2} according to Lemma (3) and the observation that $\delta_N \leq \delta_{N,\sigma^2}$, and consider the candidate pdf $f'_{\tilde{N},N,\delta} = \alpha f_{\mathcal{N}} + (1-\alpha) f_{\tilde{N},N,\delta}$ where $0 < \alpha < 1$ and α is chosen small enough so that $D(f_{\tilde{N},N,\delta}||f_N) \leq \delta$, which is achievable via convexity of the divergence measure. By concavity of the objective function (2) we have that

$$\begin{aligned} P_{\rm ML}(f_{\tilde{N},N,\delta}) &\geq \alpha \frac{\lfloor a \rfloor}{2(\lfloor a \rfloor + 1)} + (1-\alpha) P_{\rm ML}(f_{\tilde{N},N,\delta}) \\ &> P_{\rm ML}(f_{\tilde{N},N,\delta}) \end{aligned}$$

demonstrating that $f_{\tilde{N},N,\delta}$ is suboptimal and thus providing the desired contradiction. □

L. Proof of Lemma 16

In this case, the upper bound on ML probability of detection error (2) is provided by the supremum over all pdfs $\sup_f P_{\rm ML}(f) = 1/2$. For large values of divergence tolerance, we can come arbitrarily close to this bound by taking as candidate the uniform pdf $\Pi_{\overline{x}}(x)$ for increasing values of $\overline{x}$. Take for divergence tolerance any value δ and assume by contradiction that the worst case pdf $f_{\tilde{N},N,\delta}$ exhibits

$$D(f_{\tilde{N},N,\delta}||f_N) = \delta_1 < \delta.$$

Now take $\overline{x}$ large enough so that

$$P_{\rm ML}(\Pi_{\overline{x}}) > P_{\rm ML}(f_{\tilde{N},N,\delta})$$

and consider the new candidate pdf given by

$$f_{\tilde{N},N,\delta} = \alpha \Pi_{\overline{x}} + (1-\alpha) f_{\tilde{N},N,\delta}$$

where $0 < \alpha < 1$ and α is chosen small enough so that

$$D(f_{\tilde{N},N,\delta}||f_N) \leq \delta.$$

Concavity of the objective function dictates that

$$\begin{aligned} P_{\rm ML}(f_{\tilde{N},N,\delta}) &\geq \alpha P_{\rm ML}(\Pi_{\overline{x}}) + (1-\alpha) P_{\rm ML}(f_{\tilde{N},N,\delta}) \\ &> P_{\rm ML}(f_{\tilde{N},N,\delta}) \end{aligned}$$

which leads to the desired contradiction. □

M. Derivation of Proposition 1

From Section III-A we know that the worst case noise pdf $f_{\tilde{N}}$ takes the form (30) for a $\mathcal{N}(0, \sigma^2)$ nominal pdf. Since, for fixed $\overline{x} > 1$, the divergence $D(f_{\tilde{N}}||\mathcal{N}(0, \sigma^2))$ is bounded below, it is clear that $\lim_{\delta \to 0} \overline{x} = 1$. Hence we fix δ small and narrow our focus to the case $\overline{x} = 1 + \tilde{x}$ where $\tilde{x} \to 0$. It is also clear by the same reasoning that $\lim_{\delta \to 0} c = 0$. Substituting the assumption $\lfloor \overline{x} \rfloor = 1$ into (30) leads to the relevant worst case form

$$f_{\hat{N}}(x) = \begin{cases} \dfrac{k_0}{\sqrt{2\pi}\overline{\sigma}} e^{-\frac{x^2}{2\overline{\sigma}^2}}, & |x| > 1 + \tilde{x} \\ \dfrac{k_1}{\sqrt{2\pi}\overline{\sigma}} e^{-\frac{x^2}{2\overline{\sigma}}}, & |x| < 1 - \tilde{x} \\ \dfrac{k_2 e^{-\frac{1}{2\overline{\sigma}}}}{\sqrt{2\pi}\overline{\sigma}} e^{-\frac{(x-1)^2}{2\overline{\sigma}}}, & 1 - \tilde{x} \le |x| \le 1 + \tilde{x} \end{cases} \tag{51}$$

where (31) yields

$$\begin{aligned} k_0 &= \frac{1}{2}\left\{ 2e^{-\frac{1+2\tilde{x}}{2\overline{\sigma}}} \left(\frac{1}{2} - Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right) \right. \\ &\qquad \left. + e^{-\frac{4\tilde{x}}{2\overline{\sigma}}} \left(\frac{1}{2} - Q\left(\frac{1-\tilde{x}}{\overline{\sigma}}\right)\right) + Q\left(\frac{1+\tilde{x}}{\overline{\sigma}}\right) \right\}^{-1} \\ &= 1 + \Phi_1^{k_0}\tilde{x} + \Phi_2^{k_0}\tilde{x}^2 + o(\tilde{x}^2) \end{aligned}$$

with

$$\begin{aligned} \Phi_1^{k_0} &= \frac{2}{\overline{\sigma}^2}\left(1 - 2Q\left(\frac{1}{\overline{\sigma}}\right)\right) \\ \Phi_2^{k_0} &= \frac{2}{\overline{\sigma}^4}\left(1 - 2Q\left(\frac{1}{\overline{\sigma}}\right)\right)\left(1 - 4Q\left(\frac{1}{\overline{\sigma}}\right)\right). \end{aligned}$$

We denote by $Q_{\overline{\sigma}^2}$ the quantity $Q(1/\overline{\sigma})$, noting that it represents the probability of ML detection error achieved by the variance-scaled Gaussian $f_{\bar{N}}$. The values of k_1 and k_2 are similarly given through (32) and (33) by

$$\begin{aligned} k_1 &= 1 + \Phi_1^{k_1}\tilde{x} + \Phi_2^{k_1}\tilde{x}^2 + o(\tilde{x}^2) \\ k_2 &= 1 + \Phi_1^{k_2}\tilde{x} + o(\tilde{x}) \end{aligned}$$

where

$$\begin{aligned} \Phi_1^{k_1} &= -\frac{4}{\overline{\sigma}^2} Q_{\overline{\sigma}^2} \\ \Phi_2^{k_1} &= \Phi_2^{k_0} - \frac{2}{\overline{\sigma}^2}(1 - 4Q_{\overline{\sigma}^2}) \\ \Phi_1^{k_2} &= \frac{1}{\overline{\sigma}^2}(1 - 4Q_{\overline{\sigma}^2}). \end{aligned}$$

Recalling (35), we have

$$\begin{aligned} &E[\hat{N}^2] \\ &\quad = 2(1 + \Phi_1^{k_0}\tilde{x})\left[\frac{\overline{\sigma}(1+\tilde{x})}{\sqrt{2\pi}} e^{-\frac{(1+\tilde{x})^2}{2\overline{\sigma}^2}} + \overline{\sigma}^2 Q_{\overline{\sigma}^2} \right] \\ &\qquad + 2(1 + \Phi_1^{k_1}\tilde{x})\left[\overline{\sigma}^2 \left(\frac{1}{2} - Q\left(\frac{1-\tilde{x}}{\overline{\sigma}}\right)\right) \right. \\ &\qquad \left. - \frac{\overline{\sigma}(1-\tilde{x})}{\sqrt{2\pi}} e^{-\frac{(1-\tilde{x})^2}{2\overline{\sigma}^2}} \right] + 4(1 + \Phi_1^{k_2}\tilde{x}) e^{-\frac{1}{2\overline{\sigma}^2}} \\ &\qquad \times \left[(\overline{\sigma}^2 + 1)\left(\frac{1}{2} - Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right) - \frac{\overline{\sigma}\tilde{x}}{\sqrt{2\pi}} e^{-\frac{\tilde{x}^2}{2\overline{\sigma}^2}} \right] + o(\tilde{x}) \\ &\quad = \overline{\sigma}^2 + \frac{4}{\sqrt{2\pi}\overline{\sigma}} e^{-\frac{1}{2\overline{\sigma}^2}} \tilde{x} + o(\tilde{x}). \end{aligned}$$

Recalling further that $E[\hat{N}^2] = \sigma^2$ according to Lemma 9 and that $\overline{\sigma}^2 = \sigma^2(1 + 2c\sigma^2)^{-1}$ we have

$$\sigma^2 = \frac{\overline{\sigma}^2}{1 - 2c\overline{\sigma}^2} = \overline{\sigma}^2 + 2c\overline{\sigma}^4 + o(c),$$

implying that

$$\begin{aligned} c &= \Phi_c\tilde{x} + o(\tilde{x}), \\ \Phi_c &= \frac{2e^{-(1/2\overline{\sigma}^2)}}{\sqrt{2\pi}\overline{\sigma}^5}. \end{aligned}$$

The divergence of $f_{\hat{N}}$ with respect to the nominal is given through (34) by

$$\begin{aligned} &D(f_{\hat{N}} \| \mathcal{N}(0,\sigma^2)) \\ &\quad = 2(1 + \Phi_1^{k_0}\tilde{x} + \Phi_2^{k_0}\tilde{x}^2)\log(1 + \Phi_1^{k_0}\tilde{x} + \Phi_2^{k_0}\tilde{x}^2) \\ &\qquad \cdot Q\left(\frac{1+\tilde{x}}{\overline{\sigma}}\right) + 2(1 + \Phi_1^{k_1}\tilde{x} + \Phi_2^{k_1}\tilde{x}^2) \\ &\qquad \cdot \log(1 + \Phi_1^{k_1}\tilde{x} + \Phi_2^{k_1}\tilde{x}^2) \times \left(\frac{1}{2} - Q\left(\frac{1-\tilde{x}}{\overline{\sigma}}\right)\right) \\ &\qquad + 4(1 + \Phi_1^{k_2}\tilde{x})\log(1 + \Phi_1^{k_2}\tilde{x}) \times e^{-\frac{1}{2\overline{\sigma}^2}} \\ &\qquad \cdot \left(\frac{1}{2} - Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right) - (\Phi_c)^2\tilde{x}^2\overline{\sigma}^4 + o(\tilde{x}^2) \\ &\quad = \eta_1\tilde{x} + \eta_2\tilde{x}^2 + o(\tilde{x}^2) \end{aligned}$$

where

$$\begin{aligned} \eta_1 &= 2Q_{\overline{\sigma}^2}\Phi_1^{k_0} + 2\left(\frac{1}{2} - Q_{\overline{\sigma}^2}\right)\Phi_1^{k_1} \\ &= 0 \\ \eta_2 &= 2Q_{\overline{\sigma}^2}\left(\Phi_2^{k_0} - \frac{1}{2}(\Phi_1^{k_0})^2\right) + 2(\Phi_1^{k_0})^2 Q_{\overline{\sigma}^2} \\ &\quad - 2\Phi_1^{k_0}\frac{e^{-\frac{1}{2\overline{\sigma}^2}}}{\sqrt{2\pi}\overline{\sigma}} + 2\left(\frac{1}{2} - Q_{\overline{\sigma}^2}\right)\left(\Phi_2^{k_1} - \frac{1}{2}(\Phi_1^{k_1})^2\right) \\ &\quad + 2(\Phi_1^{k_1})^2\left(\frac{1}{2} - Q_{\overline{\sigma}^2}\right) - 2\Phi_1^{k_1}\frac{e^{-\frac{1}{2\overline{\sigma}^2}}}{\sqrt{2\pi}\overline{\sigma}} \\ &\quad + 4\Phi_1^{k_2}\frac{e^{-\frac{1}{2\overline{\sigma}^2}}}{\sqrt{2\pi}\overline{\sigma}} - (\Phi_c)^2\overline{\sigma}^4 \\ &= \frac{4}{\overline{\sigma}^4}Q_{\overline{\sigma}^2}(1 - 2Q_{\overline{\sigma}^2}) - \frac{2e^{-\frac{1}{\overline{\sigma}^2}}}{\pi\overline{\sigma}^6}. \end{aligned}$$

Finally, the zero-threshold detector is an ML detector for $f_{\hat{N}}$ according to Lemma 7, implying that

$$\begin{aligned} P_{\hat{N}} &= \int_1^\infty f_{\hat{N}}(x)\,dx \\ &= (1 + \Phi_1^{k_0}\tilde{x})Q\left(\frac{1+\tilde{x}}{\overline{\sigma}}\right) + (1 + \Phi_1^{k_2}) \\ &\quad \cdot e^{-\frac{1}{2\overline{\sigma}^2}}\left(\frac{1}{2} - Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right) + o(\tilde{x}) \\ &= Q_{\overline{\sigma}^2} + \Phi_1^{k_0}Q_{\overline{\sigma}^2}\tilde{x} + o(\tilde{x}). \end{aligned} \tag{52}$$

Writing $Q_{\sigma^2} = Q(1/\sigma)$ to denote the nominal probability of ML detection error, and $\Delta\hat{P} = P_{\hat{N}} - Q_{\sigma^2}$ to denote worst case power- and divergence-constrained performance degradation, we have that

$$\begin{aligned}\Delta\hat{P} &= \Phi_1^{k_0} Q_{\overline{\sigma}^2}\tilde{x} + Q_{\overline{\sigma}^2} - Q_{\sigma^2} + o(\tilde{x}) \\ &= \left[\Phi_1^{k_0} Q_{\overline{\sigma}^2} - \Phi_c \frac{\overline{\sigma} e^{-\frac{1}{2\overline{\sigma}^2}}}{\sqrt{2\pi}}\right]\tilde{x} + o(\tilde{x}) \\ &= \left[\frac{2Q_{\overline{\sigma}^2}(1-2Q_{\overline{\sigma}^2})}{\overline{\sigma}^2} - \frac{e^{-\frac{1}{2\overline{\sigma}^2}}}{\pi\overline{\sigma}^4}\right]\tilde{x} + o(\tilde{x}).\end{aligned}$$

Setting $D(f_{\hat{N}}\|\mathcal{N}(0,\sigma^2)) = \delta$ according to Lemma 4 leads to the relation

$$\frac{\Delta\hat{P}}{\sqrt{\delta}} = \frac{\left[\frac{2Q_{\overline{\sigma}^2}(1-2Q_{\overline{\sigma}^2})}{\overline{\sigma}^2} - \frac{e^{-\frac{1}{2\overline{\sigma}^2}}}{\pi\overline{\sigma}^4}\right]\tilde{x} + o(\tilde{x})}{\sqrt{\left[\frac{4Q_{\overline{\sigma}^2}(1-2Q_{\overline{\sigma}^2})}{\overline{\sigma}^4} - \frac{2e^{-\frac{1}{\overline{\sigma}^2}}}{\pi\overline{\sigma}^6}\right]\tilde{x}^2 + o(\tilde{x}^2)}}$$

which yields asymptotically

$$\lim_{\delta\to 0}\frac{\Delta\hat{P}}{\sqrt{\delta}} = \sqrt{Q_{\overline{\sigma}^2}(1-2Q_{\overline{\sigma}^2}) - \frac{e^{-\frac{1}{\overline{\sigma}^2}}}{2\pi\overline{\sigma}^2}}.$$

Recalling that $\overline{\sigma}^2$ tends to σ^2 and hence that $Q_{\overline{\sigma}^2}$ tends to Q_{σ^2} as δ tends to zero, we obtain the governing expression

$$\lim_{\delta\to 0}\frac{\Delta\hat{P}}{\sqrt{\delta}} = \sqrt{Q_{\sigma^2}(1-2Q_{\sigma^2}) - \frac{e^{-\frac{1}{\sigma^2}}}{2\pi\sigma^2}} \tag{53}$$

for worst case power- and divergence-constrained error degradation.

N. Derivation of Proposition 2

The analysis of asymptotic behavior of divergence-constrained noise $f_{\tilde{N}}$ follows very closely that of power- and divergence-constrained noise developed for Proposition 1, with the exception that the variance-scaled nominal $f_{\overline{N}}$ is replaced with the nominal f_N due to the lack of a requirement for power considerations, equivalent to setting $c = 0$ (i.e., $\Phi_c = 0$), $\overline{\sigma}^2 = \sigma^2$, and $Q_{\overline{\sigma}^2} = Q_{\sigma^2}$. Hence, the expression for divergence is given by

$$\begin{aligned}D(f_{\tilde{N}}\|\mathcal{N}(0,\sigma^2)) &= \eta_1\tilde{x} + \eta_2\tilde{x}^2 + o(\tilde{x}^2) \\ \eta_1 &= 0 \\ \eta_2 &= \frac{4}{\sigma^4}Q_{\sigma^2}(1-2Q_{\sigma^2})\end{aligned}$$

and the detection error probability is given by

$$P_{\tilde{N}} = Q_{\sigma^2} + \frac{2}{\sigma^2}Q_{\sigma^2}(1-2Q_{\sigma^2})\tilde{x} + o(\tilde{x}).$$

Worst case divergence-constrained performance degradation $\Delta\tilde{P} = P_{\tilde{N}} - Q_{\sigma^2}$ then satisfies the relation

$$\frac{\Delta\tilde{P}}{\sqrt{\delta}} = \frac{\frac{2}{\sigma^2}Q_{\sigma^2}(1-2Q_{\sigma^2})\tilde{x} + o(\tilde{x})}{\sqrt{\frac{4}{\sigma^4}Q_{\sigma^2}(1-2Q_{\sigma^2})\tilde{x}^2 + o(\tilde{x}^2)}}$$

implying that

$$\lim_{\delta\to\infty}\frac{\Delta\tilde{P}}{\sqrt{\delta}} = \sqrt{Q_{\sigma^2}(1-2Q_{\sigma^2})}. \tag{54}$$

O. Derivation of Proposition 3

Worst case power- and divergence-constrained noise $f_{\hat{N}}$ for a Gaussian nominal takes the form (30) as obtained from (17) by direct substitution. We fix the divergence tolerance δ and consider the parametric behavior of (30) as σ^2 tends to zero. The second moment of $f_{\hat{N}}$ is given by (23). It is clear from the definitions of $A_0, A_{\lfloor\overline{x}\rfloor}$ and $A_{\lfloor\overline{x}\rfloor+1}$ associated with (22) and (23) that

$$2A_0 + 2\lfloor\overline{x}\rfloor A_{\lfloor\overline{x}\rfloor} + 2(\lfloor\overline{x}\rfloor+1)A_{\lfloor\overline{x}\rfloor+1} = 1$$

since $f_{\hat{N}}$ is a proper pdf. Since $\overline{x}\ge 1$, it is always the case that $A_0 < 1/2$ for unimodal nominals, implying that

$$\lfloor\overline{x}\rfloor A_{\lfloor\overline{x}\rfloor} + (\lfloor\overline{x}\rfloor+1)A_{\lfloor\overline{x}\rfloor+1} > 0.$$

Given that

$$\begin{aligned}E[\hat{N}^2] &\ge \frac{2(\lfloor\overline{x}\rfloor-1)\lfloor\overline{x}\rfloor+1)}{3}A_{\lfloor\overline{x}\rfloor} \\ &\quad + \frac{2\lfloor\overline{x}\rfloor(\lfloor\overline{x}\rfloor+1)(\lfloor\overline{x}\rfloor+2)}{3}A_{\lfloor\overline{x}\rfloor+1} \\ &\ge \frac{2(\lfloor\overline{x}\rfloor-1)(\lfloor\overline{x}\rfloor+1)}{3}(\lfloor\overline{x}\rfloor A_{\lfloor\overline{x}\rfloor} + (\lfloor\overline{x}\rfloor+1)A_{\lfloor\overline{x}\rfloor+1})\end{aligned}$$

according to (23), it is, therefore, necessary that $\lfloor\overline{x}\rfloor = 1$ in order for $E[\hat{N}^2]$ to tend to zero as required. While it is true, then, that $\overline{x} = 1+\tilde{x}, 0<\tilde{x}<1$ for the following analysis, it is *not* necessarily the case that $\tilde{x}$ tend to zero with decreasing σ^2, as was the case for decreasing δ in the analysis of Section IV-A.

Substituting $\overline{x} = 1+\tilde{x}$ into (30) yields the form (51) for $f_{\hat{N}}$, with parametric constants k_0, k_1, and k_2 given by (31)–(33) and second moment given by (35). Noting that the parametric constants $\tilde{x}, c, k_0, k_1$, and k_2 depend on σ^2 in the following analysis, (35) yields for second moment the expression

$$\begin{aligned}&E[\hat{N}^2] \\ &= \Bigg\{\overline{\sigma}\Bigg[\frac{\overline{x}}{\sqrt{2\pi}}e^{-\frac{(1-\tilde{x})^2}{2\overline{\sigma}^2}} + \overline{\sigma}\left(\frac{1}{2} - Q\left(\frac{1-\tilde{x}}{\overline{\sigma}}\right)\right) \\ &\qquad + \overline{\sigma}Q\left(\frac{\overline{x}}{\overline{\sigma}}\right)e^{\frac{2\tilde{x}}{\overline{\sigma}^2}} - \frac{\tilde{x}}{\sqrt{2\pi}}e^{-\frac{\tilde{x}^2}{2\overline{\sigma}^2}} - \frac{1-\tilde{x}}{\sqrt{2\pi}} \\ &\qquad \cdot e^{-\frac{(1-\tilde{x})^2}{2\overline{\sigma}^2}}\Bigg] + 2e^{-\frac{1-2\tilde{x}}{2\overline{\sigma}^2}} \\ &\quad \cdot\left[(\overline{\sigma}^2+1)\left(\frac{1}{2} - Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right)\right]\Bigg\}\Bigg\{\frac{1}{2} + Q\left(\frac{\overline{x}}{\overline{\sigma}}\right) \\ &\quad - Q\left(\frac{1-\tilde{x}}{\overline{\sigma}}\right) + 2e^{-\frac{1-2\tilde{x}}{2\overline{\sigma}^2}}\left(\frac{1}{2} - Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right)\Bigg\}^{-1}.\end{aligned} \tag{55}$$

By considering the limiting behavior of $e^{-(1-2\tilde{x}/2\overline{\sigma}^2)}$ as $\overline{\sigma}^2$ tends to zero, inspection of (55) yields for fixed $\tilde{x}$ the relation

$$\lim_{\overline{\sigma}^2\to 0}E[\hat{N}^2] = \begin{cases}0, & \tilde{x} < 1/2 \\ 2/3, & \tilde{x} = 1/2 \\ 1, & \tilde{x} > 1/2.\end{cases}$$

This result makes sense in light of the expression for worst case noise (30), which describes a symmetric pdf made up of three weighted variance-$\overline{\sigma}^2$ Gaussian truncations centered at $-1, 0$, and 1 and weighted $\mathcal{N}(0,\overline{\sigma}^2)$ Gaussian tails, the weights being determined by continuity at the points $\pm(1-\tilde{x})$

and $\pm(1+\tilde{x})$. For $\tilde{x}=1/2$, the three Gaussian truncations are equally weighted, so that the second moment of the mixture tends to $2/3+\overline{\sigma}^2$ for small $\overline{\sigma}^2$; for $\tilde{x}>1/2$, the Gaussian truncations centered at $x=\pm 1$ dominate, exhibiting second moment $1+\overline{\sigma}^2$ for small $\overline{\sigma}^2$; finally, for $\tilde{x}<1/2$, the Gaussian truncation centered at $x=0$ dominates, exhibiting second moment $\overline{\sigma}^2$. For our purposes, it is clear that $\tilde{x}<1/2$, enabling $E[\hat{N}^2]$ to tend to zero with decreasing $\overline{\sigma}^2$.

Using (55) and the observations concerning $\tilde{x}$, we have that

$$\lim_{\overline{\sigma}^2} \frac{\overline{\sigma}^2+2e^{-\frac{1-2\tilde{x}}{2\overline{\sigma}^2}}\left(1-2Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right)}{E[\hat{N}^2]}=1. \tag{56}$$

Using the relations $E[\hat{N}^2]=\sigma^2$ and $\overline{\sigma}^2=\sigma^2(1+2c\sigma^2)^{-1}$, we obtain for c the limiting expression

$$\lim_{\overline{\sigma}^2} \frac{1}{c}\frac{e^{-\frac{1-2\tilde{x}}{2\overline{\sigma}}}\left(1-2Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right)}{\overline{\sigma}^4+2\overline{\sigma}^2e^{-\frac{1-2\tilde{x}}{2\overline{\sigma}^2}}\left(1-2Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right)}=1.$$

The expression (34) for divergence reduces to

$$\begin{aligned}D(f_{\hat{N}}\|\mathcal{N}(0,\sigma^2)) &= 2k_0\log k_0 Q\left(\frac{\overline{x}}{\overline{\sigma}}\right)\\ &\quad + 2k_1\log k_1\left(\frac{1}{2}-Q\left(\frac{1-\tilde{x}}{\overline{\sigma}}\right)\right)\\ &\quad + 4k_2\log k_2 e^{-\frac{1}{2\overline{\sigma}^2}}\left(\frac{1}{2}-Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right)\\ &\quad + \frac{1}{2}\log(1+2c\sigma^2)-c\sigma^2\end{aligned}$$

where, using the previous observations regarding the limiting behavior of $\tilde{x}$

$$\lim_{\sigma^2\to 0} e^{\frac{2\tilde{x}}{\overline{\sigma}^2}}/k_0=1$$
$$\lim_{\sigma^2\to 0} k_1=1$$

and

$$\lim_{\sigma^2\to 0} e^{\frac{\tilde{x}}{\overline{\sigma}^2}}/k_2=1.$$

Recalling that $D(f_{\hat{N}}\|\mathcal{N}(0,\sigma^2))=\delta$, straightforward manipulation yields the expression at the bottom of this page.

For fixed $\delta>0$, it is clear that the term $e^{-\frac{1-2\tilde{x}}{2\overline{\sigma}^2}}/\overline{\sigma}^2$ remains bounded below, implying that $\tilde{x}$ tends to $1/2$ for decreasing $\overline{\sigma}^2$, and hence that $Q(\tilde{x}/\overline{\sigma})$ tends to zero. We thus have that

$$\lim_{\sigma^2\to 0}\frac{1}{2}\log\left[1+2\frac{e^{-\frac{1-2\tilde{x}}{2\overline{\sigma}^2}}}{\overline{\sigma}^2}\right]=\delta.$$

Finally, the expression (52) for $P_{\hat{N}}$ yields

$$\lim_{\sigma^2\to 0}\frac{e^{-\frac{1-2\tilde{x}}{\overline{\sigma}^2}}}{2P_{\hat{N}}}=1$$

which, when coupled with the previous observation and expression (56) relating $\overline{\sigma}^2$ with σ^2 yields

$$\begin{aligned}\lim_{\sigma^2\to 0}\frac{P_{\hat{N}}}{\sigma^2} &= \lim_{\sigma^2\to 0}\frac{e^{-\frac{1-2\tilde{x}}{2\overline{\sigma}^2}}}{2(\overline{\sigma}^2+2e^{-\frac{1-2\tilde{x}}{2\overline{\sigma}^2}}}\\ &= \frac{1-e^{-2\delta}}{4}.\end{aligned}$$

P. Derivation of Proposition 4

Choose a nominal pdf f_N with second moment σ^2 and support $(-a,a)$. If $a>M_{\sigma^2}=\lceil\sqrt{3\sigma^2+1}-1\rceil$, the behavior of $f_{\hat{N},N,\delta}$ is directly governed by Theorem 1. With the knowledge that the cumulative distribution $F_{\hat{N},N,\delta}$ approaches the step function $F_{\sigma^2}^{SV}$ described in Section II-A for large δ, we conclude that the pdf $f_{\hat{N},N,\delta}$ must necessarily attain asymptotically large values at the points corresponding to the jumps, leading to the conclusion that c grows without bound with growing δ in the optimal parametric expression (11). For sufficiently large c we have that

$$\begin{aligned}&f_{\bar{N}}(x+\tau_1)f_{\bar{N}}(x-\tau_1)-f_{\bar{N}}(x+\tau_2)f_{\bar{N}}(x-\tau_2)\\ &\quad = C^2[f_N(x+\tau_1)f_N(x-\tau_1)e^{-c[(x+\tau_1)^2+(x-\tau_1)^2]}\\ &\qquad - f_N(x+\tau_2)f_N(x-\tau_2)e^{-c[(x+\tau_2)^2+(x-\tau)^2]}]\\ &\quad = C^2e^{-2c(x^2+\tau_1^2)}[f_N(x+\tau_1)f_N(x-\tau_1)\\ &\qquad - f_N(x+\tau_2)f_N(x-\tau_2)e^{-2c(\tau_2^2-\tau_1^2)}]\\ &\quad > 0\end{aligned}$$

for all $x\in(-a,a)$ and $0\le\tau_1<\tau_2<\min\{1,a-|x|\}$, so that $f_{\bar{N}}$ satisfies condition (13). Furthermore, the assumption $a>M_{\sigma^2}$ guarantees activity of the divergence constraint. Hence, we know through Lemma 8 that $f_{\hat{N},N,\delta}$ takes the form (17) for large enough δ, implying that $f_{\hat{N},N\delta}$ is a 2-periodic function in an interval $[-\overline{x}_\delta,\overline{x}_\delta]$ with components composed of weighted geometric means. Outside this interval, $f_{\hat{N},N,\delta}$ is simply a weighted version of the variance-scaled nominal $f_{\bar{N}}$, the weight being determined through Lemma 6 by continuity at the point $\overline{x}_\delta$. Clearly, as $\delta\to\infty$ we have that $\lfloor\overline{x}_\delta\rfloor\to M_{\sigma^2}$ and that all mass converges to the points $-M_{\sigma^2},-M_{\sigma^2}+1,\cdots,M_{\sigma^2}$ according to Theorem 1. The worst case probability of ML detection error in this case satisfies

$$\lim_{\delta\to\infty}P_{\hat{N},N,\delta}=\frac{2K^3-2K+3\sigma^2}{2K(K+1)(2K+1)}$$
$$K=\min\left\{\text{integers } k\text{: } \frac{k(k+2)}{3}\ge\sigma^2\right\}$$

as derived directly from [13].

We now turn our attention to symmetric continuous nominals f_N with support $(-a,a)$ where $a\le M_{\sigma^2}$. In this

---

$$\lim_{\sigma^2\to 0}\frac{\frac{1}{2}\log\left[1+2\frac{e^{-\frac{1-2\tilde{x}}{2\overline{\sigma}^2}}}{\overline{\sigma}^2}\left(1-2Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right)\right]-(1-2\tilde{x})\frac{e^{-\frac{1-2\tilde{x}}{2\overline{\sigma}^2}}}{\overline{\sigma}^2}\left(1-2Q\left(\frac{\tilde{x}}{\overline{\sigma}}\right)\right)}{\delta}=1.$$

case, recall that the finite support constraint, which amounts to a constraint on noise amplitude, is asymptotically more restrictive than the power constraint, resulting in asymptotic behavior governed by Lemma 3. Hence, there exists some $\delta_{N,\sigma^2} < \infty$ for which worst case noise falls into the class $\mathcal{M}(a)$ of "picket-fence" distributions for all $\delta \geq \delta_{N,\sigma^2}$, implying that

$$\lim_{\delta\to\infty} P_{\hat{N},N,\delta} = \frac{\lfloor a\rfloor}{2(\lfloor a\rfloor + 1)}.$$

The value of δ_{N,σ^2} and the form of the corresponding worst case pdf $f_{\hat{N}}$ are determined by the optimization problem

$$f_{\hat{N}} = \arg \min_{f_{\mathcal{N}}\in\mathcal{M}(a):\, \int_{-a}^{a} x^2 f_{\mathcal{N}}(x)dx \leq \sigma^2} D(f_{\mathcal{N}}\|f_N)$$

since δ_{N,σ^2} represents the minimum divergence among all pdf's in $\mathcal{M}(a)$ satisfying the power constraint. Using the fact that $f_{\hat{N}}$ belongs to the class $\mathcal{M}(a)$, the problem can be reformulated as

$$f_{\hat{N}} = \arg\min_{f_{\mathcal{N}}} \sum_{i=0}^{\lfloor a\rfloor} \int_{\lfloor a\rfloor-\check{x}}^{\lfloor a\rfloor+\check{x}} f_{\mathcal{N}}(x) \log \frac{f_{\mathcal{N}}(x)}{f_N(x-2i)}\, dx$$

subject to the constraints

$$(\lfloor a\rfloor + 1) \int_{\lfloor a\rfloor+\check{x}}^{\lfloor a\rfloor-\check{x}} f_{\mathcal{N}}(x)\, dx = 1$$

$$\sum_{i=0}^{\lfloor a\rfloor} \int_{\lfloor a\rfloor-\check{x}}^{\lfloor a\rfloor+\check{x}} (x-2i)^2 f_{\mathcal{N}}(x)\, dx \leq \sigma^2$$

where $\check{x} = a - \lfloor a\rfloor$, which leads through a straightforward Lagrange-multiplier analysis to the optimal form

$$f_{\hat{N}}(x) = \begin{cases} K\left(\prod_{i=0}^{\lfloor a\rfloor} f_{\bar{N}}(x-2i)\right)^{\frac{1}{\lfloor a\rfloor+1}}, \\ \qquad x\in[\lfloor a\rfloor - \check{x}, \lfloor a\rfloor + \check{x}] \\ f_{\hat{N}}(x+2i), \\ \qquad x+2i \in [\lfloor a\rfloor - \check{a}, \lfloor a\rfloor + \check{x}], i = 1,2,\cdots,\lfloor a\rfloor \\ 0, \qquad \text{otherwise} \end{cases}$$

where $f_{\bar{N}}(x) = Cf_N(x)e^{-cx^2}$ is the variance-scaled nominal and K is chosen to render $f_{\hat{N}}$ a proper pdf and hence satisfies

$$K = \left\{(\lfloor a\rfloor + 1)\int_{\lfloor a\rfloor-\check{x}}^{\lfloor a\rfloor+\check{x}} \left(\prod_{i=0}^{\lfloor a\rfloor} f_{\bar{N}}(x-2i)\right)^{\frac{1}{\lfloor a\rfloor+1}} dx\right\}^{-1}.$$

The value of c is uniquely determined by the relation

$$\int_{-a}^{a} x^2 f_{\hat{N}}(x)\, dx = \sigma^2$$

yielding the expression

$$\delta_{N,\sigma^2} = (\lfloor a\rfloor + 1)\log K \int_{\lfloor a\rfloor-\check{x}}^{\lfloor a\rfloor+\check{x}} f_{\hat{N}}(x)\, dx.$$

Note that, for divergence tolerance δ_{N,σ^2}, the worst case pdf $f_{\hat{N}}$ exhibits the decomposition

$$C_{\lfloor a\rfloor+1} = [\lfloor a\rfloor - \check{x}, \lfloor a\rfloor + \check{x}]$$
$$\mathcal{C}_\infty = [M^* - 1 + \check{x}, M^* + 1 - \check{x}]$$

in the optimal expression (11) where M^* takes the value 1 if $\lfloor a\rfloor$ is even and the value 0 if $\lfloor a\rfloor$ is odd. Inspection also reveals that $f_{\hat{N}}$ is continuous when the divergence tolerance is set to δ_{N,σ^2}, formally completing Lemma 6.

Q. Derivation of Proposition 5

For symmetric continuous nominals f_N with finite support $(-a,a), a < \infty$, the resulting restriction on the support of $f_{\tilde{N},N,\delta}$ is equivalent to a constraint on noise amplitude. As discussed in Section II, worst case noise for such a constraint class, without reference to a nominal, consists of the family $\mathcal{M}(a)$ of "picket-fence" distributions, all of which exhibit the ML probability of detection error $\lfloor a\rfloor/(2(\lfloor a\rfloor+1))$. This provides an upper bound for the divergence-constrained problem, which is achieved for large enough divergence tolerance by the pulse-string pdf

$$f_a(x) = \frac{1}{\lfloor a\rfloor + 1}\sum_{i=0}^{\lfloor a\rfloor} \Pi_{\check{x}/2}(x + \lfloor a\rfloor - 2i)$$

where $\check{x} = a - \lfloor a\rfloor$.

Hence, there is some range $[\delta_N, \infty)$ of divergence tolerance δ for which $f_{\tilde{N},N,\delta}$ falls in the worst case amplitude-constrained class $\mathcal{M}(a)$, implying that

$$P_{\tilde{N},N,\delta} = \lfloor a\rfloor/(2(\lfloor a\rfloor+1)).$$

In order to determine δ_N and to characterize the actual form of $f_{\tilde{N},N,\delta}$ for the case $\delta = \delta_N$, we note that the latter satisfies the optimization problem

$$f_{\tilde{N}} = \arg \min_{f_{\mathcal{N}}\in\mathcal{M}(a)} D(f_{\mathcal{N}}\|f_N)$$

which can be rewritten using the properties of the class $\mathcal{M}(a)$ in the form

$$f_{\tilde{N}} = \arg\min_{f_{\mathcal{N}}} \sum_{i=0}^{\lfloor a\rfloor} \int_{\lfloor a\rfloor-\check{x}}^{\lfloor a\rfloor+\check{x}} f_{\mathcal{N}}(x) \log \frac{f_{\mathcal{N}}(x)}{f_N(x-2i)}\, dx$$

subject to the constraint

$$(\lfloor a\rfloor + 1) \int_{\lfloor a\rfloor-\check{x}}^{\lfloor a\rfloor-\check{x}} f_{\mathcal{N}}(x)\, dx = 1$$

leading to the optimal form

$$f_{\tilde{N}}(x) = \begin{cases} K\left(\prod_{i=0}^{\lfloor a\rfloor} f_N(x-2i)\right)^{\frac{1}{\lfloor a\rfloor+1}}, \\ \qquad x\in[\lfloor a\rfloor - \check{x}, \lfloor a\rfloor + \check{x}] \\ f_{\tilde{N}}(x+2i), \\ \qquad x+2i \in [\lfloor a\rfloor - \check{x}, \lfloor a\rfloor + \check{x}], i = 1,2,\cdots,\lfloor a\rfloor] \\ 0, \quad \text{otherwise} \end{cases}$$

where K scales $f_{\tilde{N}}$ to a proper pdf and hence satisfies

$$K = \left\{ (\lfloor a \rfloor + 1) \int_{\lfloor a \rfloor - \breve{x}}^{\lfloor a \rfloor + \breve{x}} \left(\prod_{i=0}^{\lfloor a \rfloor} f_N(x - 2i) \right)^{\frac{1}{\lfloor a \rfloor + 1}} \right\}^{-1}.$$

Finally, the value of δ_N can be shown through straightforward manipulation to satisfy

$$\delta_N = (\lfloor a \rfloor + 1) \log K \int_{\lfloor a \rfloor - \breve{x}}^{\lfloor a \rfloor + \breve{x}} f_{\tilde{N}}(x)\, dx$$

which is directly computable upon specification of the nominal f_N.

For values of δ approaching δ_N from below, $f_{\tilde{N},N,\delta}$ will approach $f_{\tilde{N},N,\delta_N}$ in distribution, while for values of δ larger than δ_N, there is a family of worst case pdf's belonging to $\mathcal{M}(a)$ and including $f_{\tilde{N},N,\delta_N}$. As noted in Lemma 12, there is an optimal realization of (28) which meets the divergence constraint with equality even for $\delta > \delta_N$, which is justified by the observation that the continuity Lemma 13 does not necessarily hold for an inactive divergence constraint; this allows for realizations of the form

$$f_{\tilde{N},N,\delta}(x) = \begin{cases} K_r \left(\prod_{i=0}^{M} f_N(x-2i) \right)^{\frac{1}{M+1}}, & x \in [M-r, M+r] \\ f_{\tilde{N},N,\delta}(x+2i), & x \in [M-r-2i, M+r-2i], i = 1, \cdots, M \\ 0, & \text{otherwise} \end{cases}$$

where $0 < r \leq \breve{x}$ and K_r is chosen to give $f_{\tilde{N},N,\delta}$ unit area. We note that by varying r continuously between 0 and $\breve{x}$, any value of $D(f_{\tilde{N}} || f_N)$ greater than δ_N can be achieved, and that the resulting discontinuities at the points $M - r - 2i$ and $M + r - 2i,\ i = 0, 1, \cdots, M$ are justified by the inactivity of the divergence constraint for the case $\delta > \delta_N$.

Now consider nominals f_N with infinite support $(a = \infty)$, for which an analysis based on amplitude-constrained noise as developed above does not apply. Instead, we make use of the unconstrained supremum of $1/2$ on binary-input probability of ML detection error over the class of all probability distributions, demonstrating a construction that achieves arbitrary tightness for sufficiently large values of divergence tolerance. The uniform pdf $\Pi_{\overline{x}}$ exhibits finite divergence with respect to any nominal with infinite support, and achieves $P_{\mathrm{ML}}(\Pi_{\overline{x}}) = (2\overline{x} - 1)/4\overline{x}$; by taking $\overline{x}$ sufficiently large, $P_{\mathrm{ML}}(\Pi_{\overline{x}})$ can be forced arbitrarily close to the upper bound $1/2$, characterizing the asymptotic behavior of $P_{\tilde{N},N,\delta}$ for such nominals. The asymptotic form of $f_{\tilde{N},N,\delta}$ is harder to characterize, since there may be several realizations of (28) that exhibit an asymptotic detection error probability approaching the value $1/2$. We do know that for nominals f_N satisfying condition (15), the worst case divergence-constrained noise pdf takes the form (29) for all values of δ according to Theorem 3, providing the desired description; it is clear that since the form (29) describes a 2-periodic function in an interval $[-\overline{x}, \overline{x}]$ for sufficiently large δ, where the size of the interval grows with δ, we have that $P_{\tilde{N},N,\delta} \to 1/2$ as $\delta \to \infty$ as desired. The lack of an expression for a limiting distribution of $f_{\tilde{N},N,\delta}$ is a direct result of the lack of an unconstrained maximizing distribution achieving ML detection error probability $1/2$.

Acknowledgment

The authors wish to thank an anonymous referee for a thorough review and helpful comments relating to the various proofs.

References

[1] I. G. Stiglitz, "Coding for a class of unknown channels," *IEEE Trans. Inform. Theory*, vol. IT-12, pp. 189–195, Apr. 1966.

[2] P. van Rooyen and F. Solms, "Maxiumum entropy investigation of the inter user interference distribution in a DS/SSMA system," in *Proc. 1995 Symp. Personal Indoor and Mobile Radio Commun.* (Toronto, Ont., Canada, 1995), pp. 1308–1312.

[3] P. A. Kullstam, "Spread spectrum performance analysis in arbitrary interference," *IEEE Trans. Commun.*, vol. COM-25, pp. 848–853, Aug. 1977.

[4] B. L. Hughes, "On nonlinear direct-sequence detectors in arbitrary power-limited interference," *IEEE Trans. Commun.*, vol. 44, pp. 733–741, June 1996.

[5] J. S. Bird and E. B. Felstead, "Antijam performance of fast frequency-hopped M-ary NCFSK—An overview," *IEEE Trans. Select. Areas Commun.*, vol. SAC-4, pp. 216–233, Mar. 1986.

[6] S. W. Houston, "Modulation techniques for communication, Part I: Tone and noise jamming performance of spread spectrum M-ary FSK and 2, 4-ary DPSK waveforms," in *Proc. NAECON Conf. Rec.*, 1975, pp. 51–58.

[7] L. J. Mason, "Error probabilities for FH/MDPSK in multitone jamming, fast Rician fading, and Gaussian noise," *IEEE Trans. Commun.*, vol. 43, pp. 545–553, Feb./Mar./Apr. 1995.

[8] M. K. Simon, "The performance of M-ary FH-DPSK in the presence of partial-band multitone jamming," *IEEE Trans. Commun.*, vol. COM-30, pp. 953–958, May 1982.

[9] G. L. Stüber, "Soft-limiter receivers for coded DS/DPSK systems," *IEEE Trans. Commun.*, vol. 38, pp. 46–53, Jan. 1990.

[10] Q. Wang, T. A. Gulliver, V. K. Bhargava, and E. B. Felstead, "Performance of fast frequency hopped noncoherent MFSK with a fixed hop rate under worst case jamming," *IEEE Trans. Commun.*, vol. 38, pp. 1786–1798, Oct. 1990.

[11] F. E. Glave, "An upper bound on the probability of error due to intersymbol interference for correlated digital signals," *IEEE Trans. Inform. Theory*, vol. IT-18, pp. 356–360, May 1972.

[12] C. R. Cahn, "Worst interference for coherent binary channel," *IEEE Trans. Inform. Theory*, vol. IT-17, pp. 209–210, Mar. 1971.

[13] S. Shamai (Shitz) and S. Verdú, "Worst case power-constrained noise for binary-input channels," *IEEE Trans. Inform. Theory*, vol. 38, pp. 1494–1511, Sept. 1992.

[14] M. V. Hedge, W. E. Stark, and D.Teneketzis, "On the capacity of channels with unknown interference," *IEEE Trans. Inform. Theory*, vol. 35, pp. 770–783, July 1989.

[15] M. A. Klimesh and W. E. Stark, "Worst case power-constrained noise for binary-input channels with varying amplitude signals," in *Proc. IEEE Int. Symp. Information Theory* (Trondheim, Norway, 1994), p. 381.

[16] S. Verdú, "Multiuser detection," in *Advances in Signal Processing—Vol. 2: Signal Detection*, H. V. Poor and J. B. Thomas, Eds. Greenwich, CT: JAI, 1993, pp. 369–410.

[17] H. V. Poor and S. Verdú, "Probability of error in MMSE multiuser detection," *IEEE Trans. Inform. Theory*, vol. 43, pp. 858–871, May 1997.

[18] A. McKellips and S. Verdú, "Worst case additive noise for binary-input channels and zero-threshold detection under constraints of power and divergence," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1256-1264, July 1997.

[19] T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.

[20] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York: Wiley, 1971, vol. 2.

[21] J. Morris, "On single-sample robust detection of known signals with additive unknown-mean amplitude-bounded random interference," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 199–209, Mar. 1980.

[22] M. Pinsker, V. Prelov, and S. Verdú, "Sensitivity of channel capacity," *IEEE Trans. Inform. Theory*, vol. 41, pp. 1877–1888, Nov. 1995.

[23] A. Wyner, "Typical sequences and all that: Entropy, pattern matching, and data compression" (1994 Shannon Lecture), *IEEE Inform. Theory Soc. Newslett.*, pp. 8–15, June 1995.

[24] H. L. Royden, *Real Analysis*, 3rd ed. London, U.K.: Macmillan, 1988.