On Conditions for Linearity of Optimal Estimation

Emrah Akyol, Kumar Viswanatha and Kenneth Rose
{eakyol, kumar, rose}@ece.ucsb.edu
Department of Electrical and Computer Engineering
University of California at Santa Barbara, CA-93106.

***Abstract*—When is optimal estimation linear? It is well-known that, in the case of a Gaussian source contaminated with Gaussian noise, a linear estimator minimizes the mean square estimation error. This paper analyzes more generally the conditions for linearity of optimal estimators. Given a noise (or source) distribution, and a specified signal to noise ratio (SNR), we derive conditions for existence and uniqueness of a source (or noise) distribution that renders the L_p norm optimal estimator linear. We then show that, if the noise and source variances are equal, then the matching source is distributed identically to the noise. Moreover, we prove that the Gaussian source-channel pair is unique in that it is the only source-channel pair for which the MSE optimal estimator is linear at more than one SNR values.**

***Index Terms*—Optimal estimation, linear estimation, source-channel matching**

I. Introduction

Consider the basic problem in estimation theory, namely, source estimation from a signal received through a channel with additive noise, given the statistics of both the source and the channel. The optimal estimator that minimizes the mean square estimation error is usually a nonlinear function of the observation [1]. A frequently exploited result in estimation theory concerns the special case of Gaussian source and Gaussian channel noise, a case in which the optimal estimator is guaranteed to be linear. An open follow-up question considers the existence of other cases exhibiting such a "coincidence", and more generally the characterization of conditions for linearity of optimal estimators for general distortion measures.

This problem also has practical importance beyond theoretical interest, mainly due to significant complexity issues in both design and operation of estimators. Specifically, the optimal estimator generally involves entire probability distributions, whereas linear estimators require only up to second-order statistics for their design. Moreover, unlike the optimal estimator which can be an arbitrarily complex function that is difficult to implement, the resulting linear estimator consists of a simple matrix-vector operation. Hence, linear estimators are more prevalent in practice, despite their suboptimal performance in general. They also represent a significant temptation to "assume" that processes are Gaussian, sometimes despite overwhelming evidence to the contrary. Results in this paper identify the cases where a linear estimator is optimal, and, hence, justify the use of linear estimators in practice without recourse to complexity arguments.

The estimation problem in general has been studied intensively in the literature. It is known that, for stable distributions (which of course include the Gaussian case), the optimal estimator is linear [2], [3], [4], [5] for any signal to noise ratios (SNR). Stable distributions are a subset of the infinitely divisible distributions which, as we show in this paper, satisfy the proposed necessary condition to have a matching distribution at any SNR level. Our main contribution to the prior works (that studied linearity at all SNR levels) focuses on the linearity of optimal estimation for L_p norm and its dependence on the SNR level. We present the optimality conditions for linear estimators given a specified SNR, and for the L_p norm. As a special case, we investigate the $p = 2$ case (mean square error) in detail. Note that a similar problem has been studied in [5], [6] for $p = 2$ without analysis of the existence of the distributions satisfying the necessary condition. We show that the necessary condition of [5], [6] is indeed a special case of our necessary and sufficient conditions, and present a detailed analysis of the MSE case.

Four results are provided on the optimality of linear estimation. First, we show that if the noise (alternatively, source) distribution satisfies certain conditions, there always exists a unique source (alternatively, noise) distribution of a given power, under which the optimal estimator is linear. We further identify conditions under which such a matching distribution does *not* exist. Secondly, we show that if the source and the noise have the same variance, they *must* be identically distributed to ensure the linearity of optimal estimator. As a third result, we show that the MSE optimal estimator converges to a linear estimator for any source and Gaussian noise at asymptotically low SNR, and vice versa, for any noise and Gaussian source at asymptotically high SNR.

Having established more general conditions for linearity of optimal estimation, one wonders in what precise sense the Gaussian case may be special. This question is answered by the fourth result. We consider the optimality of linear estimation at multiple SNR values. Let random variables X and N be the source and channel noise, respectively, and allow for scaling of either to produce varying levels of SNR. We show that if the optimal estimation is linear at more than one SNR value, then both the source X and the noise N must be Gaussian[1]. In other words, the Gaussian source-noise pair is unique in that it offers linearity of optimal estimators at multiple SNR values.

The paper is organized as follows: we present the problem formulation in Section II, the main result in Section III, the

This work is supported by the NSF under the grant CCF-0728986

[1]Of course, in this case optimal estimators are linear at all SNR levels.

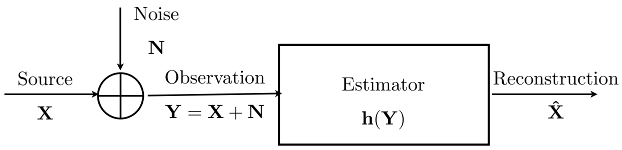

Fig. 1. The general setup of the problem

specific result for MSE in Section IV, the corollaries in Section V, comments on the vector case in Section VI and provide conclusions in Section VII.

II. Problem Formulation

A. Preliminaries and notation

We consider the problem of estimating the source X given the observation $Y = X + N$, where X and N are independent, as shown in Figure 1. Without loss of generality, we assume that X and N are scalar zero mean random variables with distributions $f_X(\cdot)$ and $f_N(\cdot)$. Their respective characteristic functions are denoted $F_X(\omega)$ and $F_N(\omega)$. A distribution $f(x)$ is said to be symmetric if it is an even function[2]: $f(x) = f(-x)\ \forall x \in R$. The SNR is $\gamma = \frac{\sigma_x^2}{\sigma_n^2}$. All distributions are constrained to have finite variance, i.e., $\sigma_x^2 < \infty, \sigma_n^2 < \infty$. All the logarithms in the paper are natural logarithms and can in general be complex.

The optimal estimator $h(\cdot)$ is the function of the observation, that minimizes the cost functional

$$J(h(\cdot)) = \mathbb{E}\left\{\Phi(X - h(Y))\right\} \tag{1}$$

for the distortion measure Φ.

B. Optimality condition for L_p norm

Re-writing (1) more explicitly,

$$J(h(\cdot)) = \int\int \Phi(x - h(y)) f_X(x) f_{Y|X}(y|x)\,dx\,dy \tag{2}$$

To obtain the necessary conditions for optimality, we apply the standard method in variational calculus [7]:

$$\frac{\partial}{\partial \epsilon} J\left[h(y) + \epsilon\eta(y)\right]\Big|_{\epsilon=0} = 0 \tag{3}$$

for all admissible variation functions $\eta(y)$. If Φ is differentiable, (3) yields

$$\int\int \Phi'(x - h(y))\eta(y) f_X(x) f_{Y|X}(y|x)\,dx\,dy = 0 \tag{4}$$

or,

$$\mathbb{E}\left\{[\Phi'(X - h(Y)]\eta(Y)\right\} = 0 \tag{5}$$

where Φ' is the derivative of Φ. This necessary condition is also sufficient for all convex Φ ($\frac{d^2\Phi}{dx^2} > 0$), in which case $\frac{\partial^2}{\partial^2 \epsilon} J\left[h(y) + \epsilon\eta(y)\right]\Big|_{\epsilon=0} > 0$, for any $\eta(y)$ variation function.

[2]Note that this definition can be generalized to symmetry about any point when one drops the assumption of zero-mean distributions

Hereafter, we will specialize our results to the case of L_p norm, i.e., $\Phi(x) = ||x||^p$ which is convex $\forall x \in \mathcal{R} - \{0\}$, ensuring the sufficiency of (5).

Note that for odd p, $\frac{d}{dx}||x||^p = p\frac{x^p}{||x||}\ \forall x \in \mathcal{R} - \{0\}$. Hence for odd p

$$\mathbb{E}\left\{\frac{[X - h(Y)]^p}{||X - h(Y)||}\eta(Y)\right\} = 0 \tag{6}$$

whereas for even p

$$\mathbb{E}\left\{[X - h(Y)]^{p-1}\eta(Y)\right\} = 0 \tag{7}$$

Note that when $\Phi(x) = x^2$, this condition reduces to the well known orthogonality condition of MSE, i.e.,

$$\mathbb{E}\left\{[(X - h(Y)]\eta(Y)\right\} = 0 \tag{8}$$

for any $\eta(\cdot)$ function. Note when $p = 2$, the optimal estimator $h(Y) = \mathbb{E}\{X|Y\}$ can be obtained from (7).

$$\int\left\{\int [x - h(y)] f_X(x) f_{Y|X}(y|x)\,dx\right\}\eta(y)\,dy = 0 \tag{9}$$

For (9) to hold for any η, the term in parenthesis should be zero, yielding $h(Y) = \mathbb{E}\{X|Y\}$, using Bayes rule.

Note that, for $p = 1$, this expression boils down to $h(Y)$ being the median, which is known as the centroid condition for L_1 norm (see e.g. [8]).

C. Optimal linear estimation for L_p norm

The linear estimator that minimizes the L_p norm is derived using linear variation functions. Plugging $\eta(y) = aY$ (for some $a \in \mathbb{R}$) in (7) and omitting some straightforward steps, we obtain the optimality condition (for even p) as

$$\mathbb{E}\left\{(X - kY)^{p-1}Y\right\} = 0 \tag{10}$$

Optimal scaling coefficient k can be found by plugging $Y = X + N$ into (10). Observe that for $p = 2$, we get the well known result $k = \frac{\gamma}{\gamma+1}$.

D. Gaussian source and channel case

We next consider the special case in which both X and N are Gaussian, $X \sim N(0, \sigma_x^2)$ and $N \sim N(0, \sigma_n^2)$. Plugging the distributions in $h(Y) = \mathbb{E}\{X|Y\}$, we obtain the well-known result

$$h(Y) = \frac{\gamma}{\gamma + 1}Y \tag{11}$$

In this case, the optimal estimator is linear at all SNR (γ) levels. Also note that it renders the estimation error $X - h(Y)$ independent of Y. It is straightforward to show that this linear estimator satisfies (6,7) and hence optimal for L_p norm. This is not a new result, it is known that optimal estimator is linear for L_p norm if both source and noise are Gaussian, see also [9].

E. Problem statement

We attempt to answer the following question: Are there other source-channel distribution pairs for which the optimal estimator turns out to be linear? More precisely, we wish to find the entire set of source and channel distributions such that $h(Y) = kY$ is the optimal estimator for some k.

III. MAIN RESULT FOR L_p NORM

In this section we derive the necessary and sufficient conditions for the linearity of optimal estimator in terms of the characteristic functions of the source and noise.

Theorem 1: For a given L_p distortion measure (p even), and given noise N with characteristic function $F_N(\omega)$, source X with characteristic function $F_X(\omega)$, the optimal estimator $h(Y)$ is linear, $h(Y) = kY$, *if and only if* the following differential equation is satisfied:

$$\sum_{m=0}^{p-1} F_X^{(m)}(\omega) F_N^{(p-1-m)}(\omega) \binom{p-1}{m} \left(\frac{k-1}{k}\right)^m = 0 \quad (12)$$

Proof: Plugging in $f_{Y|X}(y|x) = f_N(y-x)$ in (7), we obtain

$$\int [x-ky]^{p-1} f_X(x) f_N(y-x) dx = 0, \forall y \quad (13)$$

Using the binomial expansion we get

$$\sum_{m=0}^{p-1} \binom{p-1}{m} (-ky)^m \int x^{p-1-m} f_X(x) f_N(y-x) dx = 0 \quad (14)$$

Let $\otimes$ denote the convolution operator and rewrite (14) as

$$\sum_{m=0}^{p-1} \binom{p-1}{m} (-ky)^m \left[y^{p-1-m} f_X(y) \otimes f_N(y) \right] = 0 \quad (15)$$

Taking the Fourier transform (assuming the Fourier transform exists), we obtain

$$\sum_{m=0}^{p-1} \binom{p-1}{m} (-k)^m \frac{d^m}{d\omega^m} \left[\frac{d^{p-1-m}(F_X(\omega))}{d\omega^{p-1-m}} F_N(\omega) \right] = 0 \quad (16)$$

After some straightforward algebra, we obtain (12). The converse part of the theorem follows from the fact that the sufficiency of the necessary conditions (6,7) due to the convexity of the L_p norm. ■

Note that a similar condition can be obtained for odd p with the noise $F_N(\omega)$ replaced with its Hilbert transform, the details are left out due to space constraints.

IV. SPECIALIZING TO MSE

In this section, we specialize the conditions for mean square error, $p = 2$. More precisely, we wish to find the entire set of source and channel distributions such that $h(Y) = \frac{\gamma}{\gamma+1} Y$ is the optimal estimator for a given γ. Note that, this condition was derived, in another context [5], [6], albeit without consideration of important implications we focus on, including the conditions for the existence of a matching noise for a given source (and vice versa), or applications of such matching conditions. We identify the conditions for existence (and uniqueness) of a source distribution that *matches* the noise in a way that makes the optimal estimator coincide with a linear one We state the main result for MSE in the following theorem.

Theorem 2: For a given SNR level γ, and given noise N with density $f_N(n)$ and characteristic function $F_N(\omega)$, there exists a source X for which the optimal estimator is linear *if and only if* the function

$$F(\omega) = F_N(\omega)^\gamma$$

is a legitimate characteristic function. Moreover, if $F(\omega)$ is legitimate, then it is the characteristic function of the matching source, i.e. $F_X(\omega) = F(\omega)$. An equivalent theorem holds where we replace "noise" for "source" everywhere, i.e., given source and SNR level, we have a condition for existence of a matching noise.

Proof: Plugging $p = 2$ in (12) yields

$$\frac{1}{F_X(\omega)} \frac{dF_X(\omega)}{d\omega} = \gamma \frac{1}{F_N(\omega)} \frac{dF_N(\omega)}{d\omega} \quad (17)$$

or more compactly,

$$\frac{d}{d\omega} \log F_X(\omega) = \gamma \frac{d}{d\omega} \log F_N(\omega) \quad (18)$$

The solution to this differential equation is given by:

$$\log F_X(\omega) = \gamma \log F_N(\omega) + C \quad (19)$$

where C is a constant. Imposing $F_N(0) = F_X(0) = 1$, we obtain $C = 0$, hence,

$$F_X(\omega) = F_N(\omega)^\gamma \quad (20)$$

■

Hence, given a noise distribution, the necessary and sufficient condition for the existence of a matching source distribution boils down to the requirement that $F_N(\omega)^\gamma$ be a valid characteristic function. Moreover, if such a matching source exists, we have a recipe for deriving its distribution.

Bochner's theorem [4] states that a continuous $F : \mathbb{R} \to \mathbb{C}$ with $F(0) = 1$ is a characteristic function if and only if it is *positive semi-definite*. Hence, the existence of a matching source depends on the positive definiteness of $F_N(\omega)^\gamma$.

Definition: Let $f : \mathbb{R} \to \mathbb{C}$ be a complex-valued function, and $t_1, ..., t_s$ be a set of points in $\mathbb{R}$. Then f is said to be positive semi- definite (non-negative definite) if for any $t_i \in \mathbb{R}$ and $a_i \in \mathbb{C}$, $i = 1, ..., s$ we have

$$\sum_{i=1}^{s} \sum_{j=1}^{s} a_i {a_j}^* f(t_i - t_j) \geq 0 \quad (21)$$

where ${a_j}^*$ is the complex conjugate of a_j. Equivalently, we require that the $s \times s$ matrix constructed with $f(t_i - t_j)$ be positive semi-definite. If function f is positive semi-definite, its Fourier transform, $F(\omega) \geq 0, \forall \omega \in \mathbb{R}$. Hence, in the case of our candidate characteristic function, this requirement ensures that the corresponding density is indeed non-negative everywhere. We note that characterizing the entire set of $F_N(\omega)$ where $F_N(\omega)^\gamma$ is positive semi-definite may be a difficult task. Instead we illustrate with various cases of interest where $F_N(\omega)^\gamma$ is or is not positive semi-definite. Let us start with a simple but useful case.

Corollary 1: If $\gamma \in \mathbb{Z}$, a matching source distribution exists, regardless of the noise distribution.

Proof: From (20), integer γ yields to the valid characteristic function of the random variable X

$$X = \sum_{i=1}^{\gamma} N_i \tag{22}$$

where N_i are independent and identically distributed as N. ■

Let us recall the concept of infinite divisibility, which is closely related to our problem.

Definition [10]: A distribution with characteristic function $F(\omega)$ is called infinitely divisible, if for each integer $k \geq 1$, there exists a characteristic function $F_k(\omega)$ such that $F(\omega) = (F_k(\omega))^k$.

Infinitely divisible distributions have been studied extensively in probability theory [10], [11]. It is known that Poisson, exponential, and geometric distributions as well as the set of stable distributions (which includes Gaussian and Gamma distributions) are infinitely divisible. On the other hand, it is easy to see that distributions of discrete random variable with finite alphabets are not infinitely divisible.

Corollary 2: A matching source distribution exists for any positive $\gamma \in \mathbb{R}$ if $f_N(n)$ is infinitely divisible.

Proof: It is easy to show from the definition of infinite divisibility that $(F_N(\omega))^r$ is a valid characteristic function for all rational $r > 0$, using Corollary 1. Using the fact that every $\gamma \in \mathbb{R}$ is a limit of a sequence of rational numbers r_n, and by the continuity theorem [12], we conclude that $F_X(\omega) = [F_N(\omega)]^\gamma$ is a valid characteristic function. ■

However, the converse of the above corollary is not true: There can exist a matching source, even though f_N is not infinitely divisible. For example, a finite alphabet discrete random variable v is not infinitely divisible but still can be k-divisible, where $k < |V| - 1$ and $|V|$ is the cardinality of v. Hence, when $\gamma = \frac{1}{k}$, there might exist a matching source, even though noise is not infinitely divisible.

Let us now identify a case in which a matching source does not exist. When $F_N(\omega)$ is real and negative for some ω, i.e. $f_N(n)$ is symmetric but not positive semi-definite, a matching source does not exist. We state this in the form of a corollary.

Corollary 3: For $\gamma \notin \mathbb{Z}$, if $F_N(\omega) \in \mathbb{R}$ and $\exists \omega$ such that $F_N(\omega) < 0$, a matching source distribution does not exist.

Proof: We prove this corollary by contradiction. Let $F_N(\omega)$ be a valid characteristic function. Recall the orthogonality property of the optimal estimator for MSE, i.e., (8). Let $\eta(Y) = Y^m$ for $m = 1, 2, 3...M$. Plugging the best linear estimator $h(Y) = \frac{\gamma}{\gamma+1} Y$ and replacing Y with $X + N$, we obtain the condition

$$\mathbb{E}\left\{\left[X - \frac{\gamma}{\gamma+1}(X+N)\right](X+N)^m\right\} = 0 \text{ for } m = 1,..,M \tag{23}$$

Expressing $(X + N)^m$ as a binomial expansion

$$(X+N)^m = \sum_{i=0}^{m} \binom{m}{i} X^i N^{m-i} \tag{24}$$

and rearranging the terms, we obtain the $M+1$ linear equations that recursively connect all moments of $f_X(x)$ up to $M + 1$, i.e., for each $m = 1, ..., M$ we have

$$\mathbb{E}(X^{m+1}) = \gamma \mathbb{E}(N^{m+1}) + \sum_{i=0}^{m-1} A(\gamma, m, i)\mathbb{E}(N^{i+1})\mathbb{E}(X^{m-i}) \tag{25}$$

where, $A(\gamma, m, i) = \gamma\binom{m}{i} - \binom{m}{i+1}$. It follows from (25) that, if all odd moments of N are zero, then so are all odd moments of X. Hence, when the noise is symmetric, the matching source must also be symmetric. However, if $\gamma \notin \mathbb{Z}$, by (20), it follows that $F_X(\omega)$ is not real, and hence $f_X(x)$ is not symmetric. This contradiction shows that no matching source exists when $\gamma \notin \mathbb{Z}$ and noise distribution is symmetric but not positive semi-definite. ■

V. Special Cases

In this section, we return to L_p norm and investigate some special cases obtained by varying γ.

Theorem 3: Given a source and noise of equal variance, the optimal estimator is linear if and only if the noise and source distributions are identical.

Proof: For MSE, it is straightforward to see from (20) that, at $\gamma = 1$, characteristic functions must be identical. The characteristic function uniquely determines the distribution [12]. Alternatively, it can be observed directly from (12) for L_p norm that $F_N(\omega) = F_X(\omega)$ satisfies the optimality condition. ■

Theorem 4 (for MSE only): In the limit $\gamma \to 0$, the MSE optimal estimator converges to linear in probability if the channel is Gaussian, regardless of the source. Similarly, as $\gamma \to \infty$, the MSE optimal estimator converges to linear in probability if the source is Gaussian, regardless of the channel.

Proof: The proof applies the law of large numbers to (20) for MSE. For the Gaussian channel with asymptotically low SNR, (asymptotical) optimality of linear estimation can also be deduced from Eq. 91 of [13]. ■

We conjecture that this theorem also holds for L_p norm, although we currently do not have a proof.

Let us consider a setup with a given source and noise variables which may be scaled to vary the SNR γ. Can the optimal estimator be linear at different values of γ? This question is motivated by the practical setting where γ is not known in advance or may vary (e.g. in the design stage of a communication system). It is well-known that the Gaussian source-Gaussian noise pair makes the optimal estimator linear at all γ levels. Below, we show that this is the only source-channel pair whose optimal estimators are linear at multiple γ values.

Theorem 5: Let the source or channel variables be scaled to vary the SNR, γ. The L_p norm optimal estimator is linear at two different γ values γ_1 and γ_2, *if and only if* both the source and the channel noise are Gaussian.

Proof: This theorem can be proved from the set of moment equations (25). Let us say the noise is scaled by

$\alpha \in \mathbb{R}$, i.e $N_2 = \alpha N$. The relation between the moments of the original and scaled noise

$$\mathbb{E}(N_2^m) = \alpha^m \mathbb{E}(N^m) \text{ for } m = 1,..,M+1 \quad (26)$$

Also, a set of moment equations should hold for γ_1 and γ_2. For clarity, we focus on the MSE norm, but the proof for L_p norm follows the same lines. The key observation is that, as mentioned in Sec II.D, the same linear estimator is optimal for a Gaussian source-channel pair with L_p norm.

$$\mathbb{E}(X^{m+1}) = \gamma_j \mathbb{E}(N^{m+1}) + \sum_{i=0}^{m-1} A(\gamma_j, m, i)\mathbb{E}(N^{i+1})\mathbb{E}(X^{m-i}) \quad (27)$$

where $m = 1,..,M$, $j = 1,2$ and $A(\gamma, m, i) = \gamma\binom{m}{i} - \binom{m}{i+1}$. Note that every equation introduces a new variable $\mathbb{E}(X^{m+1})$, for $m = 1,..,M$, so each new equation is independent of its predecessors. Let us consider solving these equations recursively, starting from $m = 1$. At each m, we have three unknowns $(\mathbb{E}(X^{m+1}), \mathbb{E}(N^{m+1}), \mathbb{E}(N_2^{m+1}))$ that are related "linearly". Since the number of equations is equal to the number of unknowns for each m, there must exist a unique solution. We know that the moments of the Gaussian source-channel pair satisfy (27). For the Gaussian random variable, the moments uniquely determine the distribution [14], so Gaussian source and noise are the only solution. ∎

Alternate Proof: Theorem 5 can be proved, only for MSE, in an alternative way. Assume the same terminology as above. Then, $\sigma_{n_2}^2 = \alpha^2\sigma_n^2$ and $F_{N_2}(\omega) = F_N(\omega\alpha)$. Let,

$$\gamma_1 = \frac{\sigma_x^2}{\sigma_n^2},\ \gamma_2 = \frac{\sigma_x^2}{\alpha^2\sigma_n^2} \quad (28)$$

Using (20),

$$F_X(\omega) = F_N(\omega)^{\gamma_1}, F_X(\omega) = F_N(\omega\alpha)^{\gamma_2} \quad (29)$$

Taking the logarithm on both sides of (29) and plugging (28) into (29), we obtain

$$\alpha^2 = \frac{\log F_N(\alpha\omega)}{\log F_N(\omega)} \quad (30)$$

Note that (30) should be satisfied for both α and $-\alpha$ since they yield the same γ. Plugging $\alpha = -1$ in (30), we obtain $F_N(\omega) = F_N(-\omega), \forall\omega$. Using the fact that every characteristic function should be conjugate symmetric (i.e. $F_N(-\omega) = F_N^*(\omega)$), we get $F_N(\omega) \in \mathbb{R}$, $\forall\omega$. As $\log F_N(\omega)$ is $\mathbb{R} \to \mathbb{C}$, the Weierstrass theorem [15] guarantees that there is a sequence of polynomials that uniformly converges to it: $\log F_N(\omega) = k_0 + k_1\omega + k_2\omega^2 + k_3\omega^3...$, where $k_i \in \mathbb{C}$. Hence, by (30) we obtain:

$$\alpha^2 = \frac{k_0 + k_1\omega\alpha + k_2(\omega\alpha)^2 + k_3(\omega\alpha)^3...}{k_0 + k_1\omega + k_2\omega^2 + k_3\omega^3...}, \quad \forall\omega \in \mathbb{R}, \quad (31)$$

which is satisfied for all ω only if all coefficients k_i vanish, except for k_2, i.e. $\log F_N(\omega) = k_2\omega^2$, or $\log F_N(\omega) = 0 \quad \forall\omega \in \mathbb{R}$ (the solution $\alpha = 1$ is not relevant in this case). The latter is not a characteristic function, and the former is the Gaussian characteristic function, $F_N(\omega) = e^{k_2\omega^2}$ (where we use the established fact that $F_N(\omega) \in \mathbb{R}$.) Since a characteristic function determines the distribution uniquely, the Gaussian source and noise must be the only such pair. ∎

VI. Comments on the Extension to Higher Dimensions

Extension of the conditions to the vector case is nontrivial due to the fact that individual SNR values for each vector component can differ. Currently we do have the solution for this extension, but it is left out due to space constraints.

VII. Conclusion

In this paper, we derived conditions under which the optimal estimator linear for L_p norm. We identified the conditions for the existence and uniqueness of a source distribution that matches the noise in a way that ensures linearity of the optimal estimator for the special case of $p = 2$. One trivial example of this type of matching occurs for Gaussian source and Gaussian noise at all SNR levels. Another instance of matching happens when the source and noise are identically distributed where the optimal estimator is $h(Y) = \frac{1}{2}Y$. We also show that Gaussian source-channel pair is unique in that it is the only source-channel pair for which the optimal estimator is linear at more than one SNR value. Moreover, we show the asymptotical linearity of MSE optimal estimators for low SNR if the channel is Gaussian regardless of the source and vice versa, for high SNR if the source is Gaussian regardless of the channel.

References

[1] S.M. Kay, *Fundamentals of Statistical Signal Processing*, Prentice Hall PTR, 1993.

[2] V.P. Skitovic, "Linear combinations of independent random variables and the normal distribution law," *Selected Translations in Mathematical Statistics and Probability*, p. 211, 1962.

[3] S.G. Ghurye and I. Olkin, "A characterization of the multivariate normal distribution," *The Annals of Mathematical Statistics*, pp. 533–541, 1962.

[4] MM Rao and RJ Swift, *Probability Theory with Applications*, Springer, 2005.

[5] RG Laha, "On a characterization of the stable law with finite expectation," *The Annals of Mathematical Statistics*, vol. 27, no. 1, pp. 187–195, 1956.

[6] A. Balakrishnan, "On a characterization of processes for which optimal mean-square systems are of specified form," *IEEE Transactions on Information Theory*, vol. 6, no. 4, pp. 490–500, 1960.

[7] D.G. Luenberger, *Optimization by Vector Space Methods*, John Wiley & Sons Inc, 1969.

[8] A. Gersho and R.M. Gray, *Vector Quantization and Signal Compression*, Springer, 1992.

[9] S. Sherman, "Non-mean-square error criteria," *IEEE Transactions on Information Theory,*, vol. 4, no. 3, pp. 125–126, 1958.

[10] E. Lukacs, "Characteristics Functions," *Charles Griffin and Company*, 1960.

[11] F.W. Steutel and K. Van Harn, *Infinite divisibility of probability distributions on the real line*, CRC, 2003.

[12] P. Billingsley, *Probability and Measure*, John Wiley & Sons Inc, 2008.

[13] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.

[14] JA Shohat and J.D. Tamarkin, "The Problem of Moments," *New York*, 1943.

[15] R.M. Dudley, *Real Analysis and Probability*, Cambridge Univ Pr, 2002.