

On Multidimensional Optimal Estimators: Linearity Conditions

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—It is well-known that, when a multivariate Gaussian source is contaminated with Gaussian noise, a linear estimator minimizes the mean square estimation error, irrespective of the covariance matrices of both source and noise. This paper analyzes the conditions for linearity of optimal estimators for general source and noise distributions over vector spaces. Given a noise (or source) distribution, we derive conditions for existence and uniqueness of a matching source (or noise) distribution that renders the optimal estimator linear. Moreover, we establish a new characterization of the uniqueness of Gaussians: the multivariate Gaussian source-channel pair is the only pair for which the optimal estimator is linear at more than one signal-to-noise ratio.

Index Terms—Optimal estimation, linear estimation

I. INTRODUCTION

Consider a 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. A result, frequently exploited in many application fields, 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.

The linearity of regression has been well studied in the mathematical literature [1], [2]. It is well known that the set of distributions for which optimal regression is linear at all signal-to-noise ratio (SNR) values, is characterized by the stable family¹, which includes the Gaussian distribution as its only finite variance member. However, limited effort has been directed at finding more general conditions under which optimal estimators are linear. For scalar source and noise, [4] derived the linearity conditions for the mean square error (MSE) distortion measure. In our prior work [5], we derived more general conditions for scalars under L_p distortion metric and showed that they specialize to the conditions in [4]

for $p = 2$. We also analyzed the existence of distributions satisfying these conditions.

In this paper, we extend the scope to derive source-channel matching conditions for multidimensional settings under the MSE distortion criterion. The extension is nontrivial due to the dependence across the components of source and noise vectors. In an appropriate transform domain we show that linearity conditions subsume a necessary matching condition inherited from the scalar case which must hold per component of transformed noise and source. However, additional cross component constraints must be satisfied to ensure linearity of optimal estimation in the multidimensional setting.

We further analyze conditions for existence and uniqueness of distributions satisfying the matching conditions. Specifically, 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 also identify conditions under which such a matching distribution does *not* exist.

Having established more general conditions for linearity of optimal estimation, one wonders in what precise sense the multivariate Gaussian case is special in the multidimensional setting. This question is answered by the analysis of the linearity of optimal estimators at multiple SNR levels. Let random variables \mathbf{X} and \mathbf{N} be the source and channel noise, respectively, and allow for scaling of either. We show that if the optimal estimation is linear at more than one SNR level, then both the source vector \mathbf{X} and the noise vector \mathbf{N} must be multivariate Gaussian. We note that the long established characterization of the Gaussian distribution is the one that ensures linearity of optimal estimation at all SNR levels [6]. Here, we strengthen this result, as linearity of optimal estimation at only two SNR levels is sufficient to characterize Gaussians.

II. CONDITIONS FOR LINEARITY OF OPTIMAL ESTIMATION

A. Preliminaries and Notation

We consider the problem of estimating the vector source $\mathbf{X} \in \mathbb{R}^m$ given the observation $\mathbf{Y} = \mathbf{X} + \mathbf{N}$, where \mathbf{X} and $\mathbf{N} \in \mathbb{R}^m$ are independent, as shown in Figure 1. Without loss of generality, we assume that \mathbf{X} and \mathbf{N} are zero mean random variables with m -fold distributions $f_X(\cdot)$ and $f_N(\cdot)$.

This work is supported by the NSF under the grants CCF-0728986 and CCF-1016861

¹A distribution is called stable if for independent identically distributed X_1, X_2, X ; for any constants a, b ; the random variable $aX_1 + bX_2$ has the same distribution as $cX + d$ for some constants c and d [3].

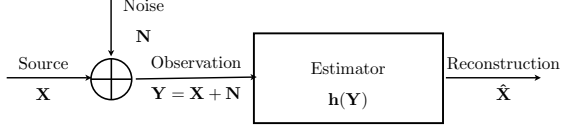


Fig. 1. The general setup of the problem

Their respective characteristic functions are denoted $F_X(\boldsymbol{\omega})$ and $F_N(\boldsymbol{\omega})$. Let $\mathbf{R}_X = \mathbb{E}\{\mathbf{X}\mathbf{X}^T\}$, $\mathbf{R}_N = \mathbb{E}\{\mathbf{N}\mathbf{N}^T\}$ be the covariance matrices of \mathbf{X} and \mathbf{N} respectively. Let \mathbf{U} be the eigenmatrix of $\mathbf{R}_X\mathbf{R}_N^{-1}$ whose eigenvalues $\lambda_1, \dots, \lambda_m$ form the diagonal elements of the diagonal matrix Λ :

$$\mathbf{R}_X\mathbf{R}_N^{-1} = \mathbf{U}\Lambda\mathbf{U}^{-1} \quad (1)$$

B. Main Result

We are looking for conditions on $F_X(\boldsymbol{\omega})$ and $F_N(\boldsymbol{\omega})$ such that $\mathbf{h}(\mathbf{Y}) = \mathbf{K}\mathbf{Y}$ where $\mathbf{K} = \mathbf{R}_X(\mathbf{R}_X + \mathbf{R}_N)^{-1}$ minimizes the estimation error $\mathbb{E}\{\|\mathbf{X} - \mathbf{h}(\mathbf{Y})\|^2\}$. Let us write the MSE optimal estimator for the vector case using Bayes' rule and the fact that \mathbf{X} and \mathbf{N} are independent.

$$\mathbf{h}(\mathbf{y}) = \frac{\int \mathbf{x}f_X(\mathbf{x})f_N(\mathbf{y} - \mathbf{x}) d\mathbf{x}}{\int f_X(\mathbf{x})f_N(\mathbf{y} - \mathbf{x}) d\mathbf{x}} \quad (2)$$

Plugging the linear form, $\mathbf{h}(\mathbf{y}) = \mathbf{K}\mathbf{y}$, we obtain,

$$\mathbf{K}\mathbf{y} \int f_X(\mathbf{x})f_N(\mathbf{y} - \mathbf{x}) d\mathbf{x} = \int \mathbf{x}f_X(\mathbf{x})f_N(\mathbf{y} - \mathbf{x}) d\mathbf{x} \quad (3)$$

Expressing the integrals as m-fold convolutions, we get

$$\mathbf{K}\mathbf{y} [f_X(\mathbf{y}) \otimes f_N(\mathbf{y})] = [\mathbf{y}f_X(\mathbf{y})] \otimes f_N(\mathbf{y}) \quad (4)$$

Taking Fourier transform of both sides,

$$j\mathbf{K}\nabla [F_X(\boldsymbol{\omega})F_N(\boldsymbol{\omega})] = jF_N(\boldsymbol{\omega})\nabla F_X(\boldsymbol{\omega}) \quad (5)$$

Arranging the terms, we get

$$(\mathbf{I} - \mathbf{K}) \frac{1}{F_X(\boldsymbol{\omega})} \nabla F_X(\boldsymbol{\omega}) = \mathbf{K} \frac{1}{F_N(\boldsymbol{\omega})} \nabla F_N(\boldsymbol{\omega}) \quad (6)$$

Using $\nabla \log F_X(\boldsymbol{\omega}) = \frac{1}{F_X(\boldsymbol{\omega})} \nabla F_X(\boldsymbol{\omega})$, and substituting $\mathbf{K} = \mathbf{R}_X(\mathbf{R}_X + \mathbf{R}_N)^{-1}$, we obtain

$$\nabla \log F_X(\boldsymbol{\omega}) = \mathbf{R}_X\mathbf{R}_N^{-1} \nabla \log F_N(\boldsymbol{\omega}) \quad (7)$$

Using the eigen decomposition in (1), we obtain

$$\mathbf{U}^{-1} \nabla \log F_X(\boldsymbol{\omega}) = \Lambda \mathbf{U}^{-1} \nabla \log F_N(\boldsymbol{\omega}) \quad (8)$$

as a necessary and sufficient condition for linearity of optimal estimation.

We will make use of the following auxiliary lemma in matrix analysis, which is stated without proof due to space constraints.

Lemma 1: Given a function $f : \mathbb{R}^n \rightarrow \mathbb{R}$, matrix $\mathbf{A} \in \mathbb{R}^{n \times m}$ and vector $\mathbf{x} \in \mathbb{R}^m$

$$\nabla_{\mathbf{x}} f(\mathbf{A}\mathbf{x}) = \mathbf{A}^T \nabla f(\mathbf{A}\mathbf{x}) \quad (9)$$

Next, we state the main theorem.

Theorem 1: Let the characteristic functions of the transformed source and noise ($\mathbf{U}\mathbf{X}$ and $\mathbf{U}\mathbf{N}$) be $F_{UX}(\boldsymbol{\omega})$ and $F_{UN}(\boldsymbol{\omega})$. The necessary and sufficient condition for linearity of optimal estimation is:

$$\frac{\partial \log F_{UX}(\boldsymbol{\omega})}{\partial \omega_i} = \lambda_i \frac{\partial \log F_{UN}(\boldsymbol{\omega})}{\partial \omega_i}, 1 \leq i \leq m \quad (10)$$

Proof: Using Lemma 1, we can rewrite (8) as

$$\nabla_{\boldsymbol{\omega}} \log F_X(\mathbf{U}^{-1}\boldsymbol{\omega}) = \Lambda \nabla_{\boldsymbol{\omega}} \log F_N(\mathbf{U}^{-1}\boldsymbol{\omega}) \quad (11)$$

Note that the characteristic functions of the source and noise after transformation can be written in terms of the known characteristic functions $F_X(\boldsymbol{\omega})$ and $F_N(\boldsymbol{\omega})$, specifically $F_{UX}(\boldsymbol{\omega}) = \det(\mathbf{U})F_X(\mathbf{U}^{-1}\boldsymbol{\omega})$ and $F_{UN}(\boldsymbol{\omega}) = \det(\mathbf{U})F_N(\mathbf{U}^{-1}\boldsymbol{\omega})$, where $\det(\cdot)$ denotes the determinant. The necessary and sufficient condition of (11) can thus be converted to the set of m scalar differential equations of (10). ■

Further insight into the above necessary and sufficient condition is provided via the following corollaries.

Corollary 1: Let $F_{UX_i}(\omega)$ and $F_{UN_i}(\omega)$ be the marginal characteristic functions of the transform coefficients $[\mathbf{UX}]_i, [\mathbf{UN}]_i$. A necessary condition for linearity of optimal estimation is:

$$F_{UX_i}(\omega) = F_{UN_i}^{\lambda_i}(\omega), 1 \leq i \leq m \quad (12)$$

Proof: Integrating both sides of (10) over all $\omega_j, j \neq i$, yields the following set of differential equations

$$\frac{\partial \log F_{UX_i}(\omega)}{\partial \omega} = \lambda_i \frac{\partial \log F_{UN_i}(\omega)}{\partial \omega}, 1 \leq i \leq m \quad (13)$$

which, given the boundary conditions $F_{UX_i}(0) = F_{UN_i}(0) = 1$, leads to the solution specified in (12) as an explicit matching condition. ■

Corollary 2: A necessary condition for linearity of optimal estimation is that one of the following holds for every pair $i, j, 1 \leq i, j \leq m$:

- i) $\lambda_i = \lambda_j$
- ii) $[\mathbf{UX}]_i$ is independent of $[\mathbf{UX}]_j$ and $[\mathbf{UN}]_i$ is independent of $[\mathbf{UN}]_j$.

Proof: Let us rewrite (10) explicitly for the i^{th} and j^{th} coefficients.

$$\frac{\partial \log F_{UX}(\boldsymbol{\omega})}{\partial \omega_i} = \lambda_i \frac{\partial \log F_{UN}(\boldsymbol{\omega})}{\partial \omega_i} \quad (14)$$

$$\frac{\partial \log F_{UX}(\boldsymbol{\omega})}{\partial \omega_j} = \lambda_j \frac{\partial \log F_{UN}(\boldsymbol{\omega})}{\partial \omega_j} \quad (15)$$

Take the partial derivative of both sides of (14) with respect to ω_j and both sides of (15) with respect to ω_i , to obtain the following:

$$\frac{\partial^2 \log F_{UX}(\boldsymbol{\omega})}{\partial \omega_i \partial \omega_j} = \lambda_i \frac{\partial^2 \log F_{UN}(\boldsymbol{\omega})}{\partial \omega_i \partial \omega_j} \quad (16)$$

$$\frac{\partial^2 \log F_{UX}(\boldsymbol{\omega})}{\partial \omega_i \partial \omega_j} = \lambda_j \frac{\partial^2 \log F_{UN}(\boldsymbol{\omega})}{\partial \omega_i \partial \omega_j} \quad (17)$$

There are only two ways to simultaneously satisfy (16) and (17): i) $\lambda_i = \lambda_j$ ii) the second order derivatives vanish, i.e., $\frac{\partial^2 \log F_{UX}(\boldsymbol{\omega})}{\partial \omega_i \partial \omega_j} = \frac{\partial^2 \log F_{UN}(\boldsymbol{\omega})}{\partial \omega_i \partial \omega_j} = 0$ which means independence of the i^{th} and j^{th} transform coefficients of source X and similarly of noise N . ■

Corollary 3: If the necessary condition of Corollary 1 is satisfied, then a sufficient condition for linearity of optimal estimation is that \mathbf{U} generates independent coefficients for both X and N .

Proof: Independence of the transform coefficients implies that the joint characteristic function is the product of the marginals:

$$F_{UX}(\boldsymbol{\omega}) = \prod_{i=1}^m F_{UX_i}(w_i), \quad F_{UN}(\boldsymbol{\omega}) = \prod_{i=1}^m F_{UN_i}(w_i) \quad (18)$$

Plugging (18) into the necessary and sufficient condition (10) of Theorem 1, it is straightforward to show that (12), the necessary condition of Corollary 1, is now both necessary and sufficient. ■

While the condition in Corollary 3 involves independence of transform coefficients, the weaker property of uncorrelatedness is already guaranteed by transform \mathbf{U} . The matrix \mathbf{U} diagonalizes both \mathbf{R}_X and \mathbf{R}_N . We formalize this in the following lemma:

Lemma 2: Transform \mathbf{U} decorrelates both source and noise: both $\mathbf{U}\mathbf{R}_X\mathbf{U}^T$ and $\mathbf{U}\mathbf{R}_N\mathbf{U}^T$ are diagonal matrices.

Proof: Since both \mathbf{R}_X and \mathbf{R}_N are, by definition, positive definite matrices, there exists a matrix \mathbf{S} that simultaneously diagonalizes \mathbf{R}_X and whitens \mathbf{R}_N , i.e., $\mathbf{S}\mathbf{R}_X\mathbf{S}^T = \boldsymbol{\Lambda}_X$ and $\mathbf{S}\mathbf{R}_N\mathbf{S}^T = \mathbf{I}$ where $\boldsymbol{\Lambda}_X$ is diagonal and \mathbf{I} is the identity matrix [7]. Hence, \mathbf{R}_X and \mathbf{R}_N can be expressed as the following:

$$\mathbf{R}_X = \mathbf{S}^{-1}\boldsymbol{\Lambda}_X\mathbf{S}^{-T}, \quad \mathbf{R}_N = \mathbf{S}^{-1}\mathbf{S}^{-T} \quad (19)$$

Plugging the above into (1) we get $\mathbf{U} = \boldsymbol{\Lambda}_U\mathbf{S}$, where $\boldsymbol{\Lambda}_U$ is diagonal. Substituting \mathbf{U} in $\mathbf{U}\mathbf{R}_X\mathbf{U}^T$ and $\mathbf{U}\mathbf{R}_N\mathbf{U}^T$, we get:

$$\mathbf{U}\mathbf{R}_X\mathbf{U}^T = \boldsymbol{\Lambda}_U\boldsymbol{\Lambda}_X\boldsymbol{\Lambda}_U^T, \quad \mathbf{U}\mathbf{R}_N\mathbf{U}^T = \boldsymbol{\Lambda}_U\boldsymbol{\Lambda}_U^T \quad (20)$$

The product of diagonal matrices is also diagonal. ■

As an example where the optimal estimator is known to be linear, consider the Gaussian multivariate case. Note that the Gaussian source-channel pair satisfies the scalar matching condition for any SNR, it satisfies the necessary condition of Corollary 1. As a linear transform preserves joint Gaussianity in the transform domain, \mathbf{U} generates jointly Gaussian and uncorrelated coefficients which are therefore independent, satisfying the conditions of Corollary 3.

Another, perhaps surprising, example where the optimal estimator is linear involves identically distributed source \mathbf{X} and noise \mathbf{N} . In this case, the linear estimator is optimal *irrespective of the distribution* shared by the source and noise. It is straightforward to show that the necessary and sufficient condition of either Theorem 1 or (7), is satisfied by $F_X(\boldsymbol{\omega}) = F_N(\boldsymbol{\omega})$.

C. Specializing to Scalars: The Matching Condition

In this section, we specialize the conditions in Theorem 1 to scalars. This result has appeared in our recent paper [5]; we include this special case for the sake of completeness. 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 , where $\gamma = \frac{\mathbb{E}(X^2)}{\mathbb{E}(N^2)}$ denotes SNR.

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

$$F(\omega) = F_N^\gamma(\omega) \quad (21)$$

is a legitimate characteristic function. Moreover, if $F(\omega)$ is legitimate, then it is the characteristic function of the matching source, i.e. $F_X(\omega) = F(\omega)$.

1) *Existence of a Matching Source for a Given Noise:* Bochner's theorem [3] states that a continuous function $F: \mathbb{R} \rightarrow \mathbb{C}$ with $F(0) = 1$ is a characteristic function if and only if it is *positive semi-definite*. We illustrate with various cases of interest where $F_N^\gamma(\omega)$ is, or is not, positive semi-definite. The proofs for the following corollaries can be found in our prior paper [5].

Corollary 4: If SNR $\gamma \in \mathbb{Z}$, a matching source distribution exists, regardless of the noise distribution.

Next, we recall the concept of infinite divisibility, which is closely related to our problem. 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 $F(\omega) = F_k^k(\omega)$.

Corollary 5: A matching source distribution exists for all $\gamma \in \mathbb{R}^+$ if and only if $f_N(n)$ is infinitely divisible.

Next, we identify a case where a matching source does not exist.

Corollary 6: For $\gamma \notin \mathbb{Z}$, if $F_N(\omega)$ is analytic² and real, and $F_N(\omega) < 0$ for $\exists \omega$, a matching source distribution does not exist.

Let us provide a commonly used example distribution that the above corollary applies to: uniform distribution over interval $[-a, a]$. In this case, $f_N(n)$ is symmetric with an analytic characteristic function, but not positive semi-definite. The corollary states that, except for integer SNR, the optimal estimator is strictly nonlinear for an additive uniform noise.

2) *Uniqueness of a Matching Source for a Given Noise:* Note (21) may have multiple solutions due to multiplicity of complex roots. The following corollary establishes that for a large set of source (or noise) distributions, the matching noise (or source) is unique.

Corollary 7: If $F_N(\omega)$ is analytic, then the matching $F_X(\omega)$, if it exists, is unique.

²A characteristic function $F(\omega)$ is analytic *if and only if* F has finite moments of all orders and there exists a finite β such that $\mathbb{E}\{|X^k|\} \leq k!\beta^k, \forall k \in \mathbb{Z}^+$. A characteristic function $F(\omega)$ is analytic *if and only if* the moments $\mathbb{E}\{|X^k|\}$ uniquely characterize X which in general is not the case, see eg. [8].

Proof: Recall the orthogonality property:

$$\mathbb{E}\{[(X - h(Y))\eta(Y)]\} = 0, \forall \eta(\cdot) \quad (22)$$

Let $\eta(Y) = Y^m$ for $m = 1, 2, 3 \dots M$. Plugging the best linear estimator $h(Y) = \frac{\gamma}{\gamma+1}Y$ and $Y = X + N$, we obtain:

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

Applying the binomial expansion for $(X+N)^m$ and rearranging the terms, we obtain $M+1$ linear equations that recursively connect the $M+1$ moments of X , i.e., for $m = 1, \dots, M$,

$$\mathbb{E}(X^{m+1}) = \gamma\mathbb{E}(N^{m+1}) + \sum_{i=0}^{m-1} A(\gamma, m, i)\mathbb{E}(N^{i+1})\mathbb{E}(X^{m-i}) \quad (24)$$

where, $A(\gamma, m, i) = \gamma\binom{m}{i} - \binom{m}{i+1}$. Note that N has finite moments and $\mathbb{E}\{|N^k|\} \leq k!\beta^k$ for all k as required to possess an analytic characteristic function. From (24), by induction, we can show that all moments of X exist and are bounded by $\mathbb{E}\{|X^k|\} \leq k!(\max(\gamma, 1)\beta)^k$. This condition is sufficient to show that X also has an analytic characteristic function.

Note that every equation introduces a new variable $\mathbb{E}(X^{m+1})$, for $m = 1, \dots, M$, 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 one unknown ($\mathbb{E}(X^{m+1})$) in a ‘‘linear’’ equation in terms of the unknown. Since the number of equations is equal to the number of unknowns for each m , and the equations are linear in terms of the unknown, there must exist a unique moment sequence that solves (24). The moments fully characterize X since it has been established that its characteristic function is analytic, hence, X , (if it exists) is unique. ■

III. A NEW CHARACTERIZATION OF GAUSSIAN DISTRIBUTION

It is well known that linearity of regression for all SNR levels characterizes the stable family of distributions, which includes Gaussian as the only finite variance member [4], [6]. We consider the problem setup where linearity of multidimensional optimal estimation is guaranteed for two different noise powers obtained by simple scaling $\mathbf{N}_2 = \alpha\mathbf{N}_1$, $\alpha \in \mathbb{R}^+$.

Theorem 3: Let the noise vector be scaled by a scalar $\alpha \in \mathbb{R}^+$ to vary the noise covariance \mathbf{R}_{N_1} . The optimal multidimensional estimator is linear at \mathbf{R}_{N_1} and $\mathbf{R}_{N_2} = \alpha^2\mathbf{R}_{N_1}$, if and only if both the source and the noise are Gaussian.

Proof: Recall that the necessary condition in Corollary 1 must hold at both SNR levels for each component, because the transform \mathbf{U} is unchanged, only component SNR levels are scaled. Hence, it is enough to analyze the scalar case. Let N_1 and N_2 denote the noise random variables with characteristic functions $F_{N_1}(\omega)$, $F_{N_2}(\omega)$ respectively, then $N_2 = \alpha N_1$ and hence $F_{N_2}(\omega) = F_{N_1}(\omega\alpha)$. Let,

$$\gamma_1 = \frac{\sigma_x^2}{\sigma_{n_1}^2}, \quad \gamma_2 = \frac{\sigma_x^2}{\alpha^2\sigma_{n_1}^2} \quad (25)$$

Using (21),

$$F_X(\omega) = F_{N_1}^{\gamma_1}(\omega), \quad F_X(\omega) = F_{N_1}^{\gamma_2}(\omega\alpha) \quad (26)$$

Hence,

$$F_{N_1}^{\gamma_1}(\omega) = F_{N_1}^{\gamma_2}(\omega\alpha) \quad (27)$$

Applying (25) on both sides of (27), we obtain

$$\alpha^2 = \frac{\log F_{N_1}(\alpha\omega)}{\log F_{N_1}(\omega)} \quad (28)$$

Note that (28) should be satisfied for both α and $-\alpha$ since they yield the same γ . Plugging $\alpha = -1$ in (28), we obtain $F_{N_1}(\omega) = F_{N_1}(-\omega)$, $\forall \omega$. Using the fact that the characteristic function is conjugate symmetric (i.e., $F_{N_1}(-\omega) = F_{N_1}^*(\omega)$), we get $F_{N_1}(\omega) \in \mathbb{R}$, $\forall \omega$. As $\log F_{N_1}(\omega)$ is $\mathbb{R} \rightarrow \mathbb{C}$, the Weierstrass theorem [10] guarantees that there is a sequence of polynomials that uniformly converges to it: $\log F_{N_1}(\omega) = \sum_{i=0}^{\infty} k_i \omega^i$, where $k_i \in \mathbb{C}$. Hence, by (28) we obtain:

$$\alpha^2 = \frac{\sum_{i=0}^{\infty} k_i (\omega\alpha)^i}{\sum_{i=0}^{\infty} k_i \omega^i}, \quad \forall \omega \in \mathbb{R}, \quad (29)$$

which is satisfied for all ω only if all coefficients k_i vanish, except for k_2 , i.e. $\log F_{N_1}(\omega) = k_2 \omega^2$, or $\log F_{N_1}(\omega) = 0 \quad \forall \omega \in \mathbb{R}$ (the solution $\alpha = 1$ is of no interest). The latter is not a characteristic function, and the former is the Gaussian characteristic function, $F_{N_1}(\omega) = e^{k_2 \omega^2}$, where we use the established fact that $F_{N_1}(\omega) \in \mathbb{R}$. Since a characteristic function determines the distribution uniquely, the Gaussian source and noise must be the only such pair. ■

IV. CONCLUSION

In this paper, we considered optimal estimation in vector spaces, and derived conditions under which the optimal estimator is linear. We identified the conditions for the existence and uniqueness of a source that matches the noise in a way that ensures linearity of the optimal estimator. We also showed that multivariate Gaussian source-channel is the only pair for which the optimal estimator is linear at multiple SNR.

REFERENCES

- [1] C. Rothschild and E. Mourier, ‘‘Sur les lois de probabilit e   regression lin aire et  cart type li  constant,’’ *Comptes Rendus*, vol. 225, 1947.
- [2] H.V. Allen, ‘‘A theorem concerning the linearity of regression,’’ *Statistical Research Memoirs*, vol. 2, pp. 60–68, 1938.
- [3] P. Billingsley, *Probability and Measure*, John Wiley & Sons Inc, 2008.
- [4] R.G. Laha, ‘‘On a characterization of the stable law with finite expectation,’’ *The Annals of Mathematical Statistics*, vol. 27, no. 1, pp. 187–195, 1956.
- [5] E. Akyol, K. Viswanatha, and K. Rose, ‘‘On conditions for linearity of optimal estimation,’’ in *Proceedings of IEEE Information Theory Workshop*, 2010.
- [6] C.R. Rao, ‘‘On some characterisations of the normal law,’’ *The Indian Journal of Statistics, Series A*, vol. 29, no. 1, pp. 1–14, 1967.
- [7] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.
- [8] J.A. Shohat and J.D. Tamarkin, *The Problem of Moments*, American Mathematical Society, 1943.
- [9] C.R. Rao, ‘‘Note on a problem of Ragnar Frisch,’’ *Econometrica, Journal of the Econometric Society*, vol. 15, no. 3, pp. 245–249, 1947.
- [10] R.M. Dudley, *Real Analysis and Probability*, Cambridge Univ Press, 2002.