Linearity Conditions for Optimal Estimation from Multiple Noisy Measurements

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

***Abstract*—Source estimation from several noisy observations, each of which are corrupted with additive independent noise is arguably among the most fundamental problems in estimation theory. Estimating multiple independent sources, corrupted by identical noise realization is important in various communication applications. It is well-known for both cases that when all sources and noises are Gaussian, linear estimator minimizes the mean square estimation error. This paper analyzes the conditions for linearity of optimal estimation in these two settings, for general source and noise distributions and distortion measures. Specifically, we show that these settings depart from the single source-single channel setting in that Gaussianity of all system components is necessary to render the L_p optimal estimator linear at a given signal-to-noise ratio (SNR). Moreover, we show for both settings that at asymptotically high SNR, for Gaussian sources the optimal estimator converges to linear, irrespective of the distribution of the noises; similarly, at low SNR, it is asymptotically linear for Gaussian noises regardless of the sources.**

***Index Terms*—Optimal estimation, linear estimation**

I. Introduction

The linearity of regression/estimation is an important and well studied problem [1], [2], [3], [4] at the core of estimation theory. It is well known that the set of distributions for which optimal regression (i.e., estimation) is linear at all signal-to-noise ratio (SNR) values, is characterized by the stable family[1], 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 a given particular SNR level.

In our prior work [6], we addressed the problem of linearity of optimal estimation of a single source from a single noisy observation, from now on referred as single source-single noise (SSSN) setting, depicted in Fig1. In particular, we derived the necessary and sufficient condition for linearity of optimal estimation with respect to the L_p distortion metric, at a given SNR. There are several implications of this study, one of which states that there is an infinite number of source-noise pairs that guarantee linearity of optimal estimation at a given SNR. One trivial example is identically distributed source and noise, regardless of the distribution, where the L_p optimal estimator is shown to be linear. This simple fact contradicts the popular belief that "only" the Gaussian source-noise pair makes the optimal estimator linear. Another important result from our prior studies is that as at asymptotically high SNR, the optimal estimator is linear for a Gaussian source, regardless of the noise distribution; and dually, for a Gaussian noise, at low SNR the optimal estimator converges to linear, irrespective of the source. We then extended the scope to derive source-channel matching conditions for the vector setting, where source and noise are vectors of the same dimension [7], [6]; in particular, we derived a necessary and sufficient condition for MSE optimality of linear estimator in the vectors setting.

In this paper, we consider two important settings that may formally be regarded as extreme special cases of the vector problem[2]. The first one is the "single source-multiple measurements" (SSMM) setting where a source is estimated by multiple observations, which are corrupted with additive independent noise. This setting is not only the most basic estimation problem in the experimental sciences but also represents a common operation in most signal processing applications, such as denoising or compressed sensing. The second one, the "multiple sources-single noise" (MSSN) setting, concerns multiple independent sources which are corrupted by the same noise. It is common in communication problems that involves estimating sources that are transmitted over a shared medium, such as sensor networks.

As our main results, we prove that the optimal estimator is linear *if and only if* all source(s) and noise(s) are Gaussian, for each of these practically relevant settings. This fact is in sharp contrast with earlier results for the SSSN setting [6]: there exists an infinite number of source-noise pairs that render the optimal estimator linear in the SSSN setting. This sharp contrast raises the question whether the other results derived within the SSSN setting generalize to these important practical settings. As we show in the paper, the results pertaining the linearity of optimal estimation at SNR asymptotics remain valid. Specifically, we prove for both settings that at asymptotically high SNR, the optimal estimator for Gaussian source(s) converges to linear, irrespective of the noise distribution; and at asymptotically low SNR, the optimal estimator is linear for

This work is supported by the NSF under the grants CCF-0728986, CCF-1016861 and CCF 1118075.

[1]A 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 [5].

[2]Note that the solution in [6] does not provide an explicit answer to these important special cases, since it implicitly assumes that both source and noise covariance matrices are invertible, while here, for both settings of interest, one of these matrices is singular.

978-1-4673-0183-1/12/$31.00 ©2012 IEEE

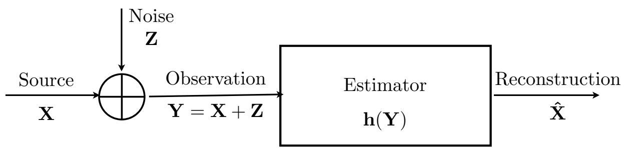

Fig. 1. The "single source-single noise" setting

Gaussian noise(s), regardless of the source(s).

This paper is organized as follows. We present a review of the SSSN setting in Section II, results within the SSMM setting in Section III, the MSSN setting in Section IV, the discussions in Section V.

II. Review of Prior Results

A. Preliminaries and Notation

Without loss of generality, we assume that source X and noise Z are zero mean random variables distributed according to $f_X(\cdot)$ and $f_Z(\cdot)$, with the respective characteristic functions $F_X(\omega)$ and $F_Z(\omega)$. The SNR is $\gamma = \frac{\sigma_x^2}{\sigma_z^2}$, where $\sigma_x^2 = \mathbb{E}\{X^2\}$ and $\sigma_z^2 = \mathbb{E}\{Z^2\}$. In any results involving the L_p norm, all random variables are assumed to have finite p^{th} order moments, e.g., in the case of MSE we assume finite variances, $\sigma_x^2 < \infty, \sigma_z^2 < \infty$. $f'(\cdot)$ denotes the derivative of the function f with respect to its argument. All logarithms in the paper are natural and may in general be complex. An estimator $h(\cdot)$ is a function of the observations, $Y_1, Y_2, ..Y_L$, and is said to be optimal if it minimizes the cost functional

$$\mathbb{E}\left\{[X - h(Y_1, Y_2.., Y_L)]^p\right\} \tag{1}$$

where p is even[3] and natural, i.e., $p = 2\rho,\ \rho \in \mathbb{N}$.

B. Single Source-Single Noise Setting

Consider the problem of estimating source X given observation $Y = X + Z$, where X and Z are independent, as shown in Fig. 1. In [6], we derived the necessary and sufficient condition for linearity of optimal estimation. We reproduce this result here in the following theorem.

Theorem ([6]). *Consider the problem setting depicted in Fig.1. Given SNR level γ, noise Z with characteristic function $F_Z(\omega)$, there exists a source X for which the MSE optimal estimator is linear* if and only if *the following is satisfied:*

$$F_X(\omega) = F_Z^{\gamma}(\omega) \tag{2}$$

This result has several important implications, see [6] for details. Two of these results, however, are of particular interest here.

Result 1: There are infinitely many source and noise pairs that yield optimality of linear estimation. For example, as long as X is identically distributed with Z, regardless of that distribution, the optimal estimator is linear, i.e., $h(Y) = Y/2$.

[3]The restriction to even p enables considerable simplification of the results, hence providing much insight and clear intuitive interpretation of the solution.

This result surprisingly remains valid for also L_p norm (see Theorem 3 in [6]).

Result 2: In the low SNR limit $\gamma \to 0$, the MSE optimal estimator is asymptotically linear if the noise is Gaussian, regardless of the source distribution. Similarly, as $\gamma \to \infty$, the MSE optimal estimator is asymptotically linear if the source is Gaussian, regardless of the noise. (see Theorem 5 in [6]).

III. Single Source- Multiple Measurements Setting

In this section, we analyze the single source-multiple measurement (SSMM) setting depicted in Fig. 2. Here, a single source X is estimated from the observations (measurements), $Y_i = X + Z_i$, $i = 1, 2, ..., L$ where Z_i's are independent noise variables. $\gamma_i = \frac{\sigma_X^2}{\sigma_{Z_i}^2}$ denote the SNR for the i^{th} observation. Let us first review the optimal linear estimation for this setting.

A. Gaussian Source-Noise Case

If $X \sim \mathcal{N}(0, \sigma_X^2)$ and $Z_i \sim \mathcal{N}(0, \sigma_{Z_i}^2)$, the MSE optimal estimator is known to be linear.

$$h(Y_1, ..., Y_L) = k_1 Y_1 + k_2 Y_2, ..., k_L Y_L \tag{3}$$

where $k_i = \frac{\gamma_i}{1 + \sum_{j=1}^{L} \gamma_j}$ which follows from standard linear estimation results.The following auxiliary lemma will be used to derive the consequent results.

Lemma 1. *A necessary condition for a function $h(Y_1, ..., Y_L)$ to be the L_p norm optimal estimator is:*

$$\mathbb{E}\{[X - h(Y_1, ..., Y_L)]^{p-1} \eta(Y_1, ..., Y_L)\} = 0 \tag{4}$$

for any $\eta(Y_1, ..., Y_L)$.

The proof follows from the proof of Lemma 1 in [6]. It follows from Lemma 1 that linear estimator, (3), is also optimal for L_p norm when $X \sim \mathcal{N}(0, \sigma_X^2)$ and $Z_i \sim \mathcal{N}(0, \sigma_{Z_i}^2)$ since this estimator renders the estimation error, $X - h(Y_1, ..., Y_L)$ independent of the observations $Y_1, ..., Y_L$, hence automatically satisfying (4).

B. Main Result

Next, we focus on the question: Are there any source and noise densities such that the optimal estimation is linear, other than the all Gaussian setting. The following theorem answers this question.

Theorem 1. *The MSE optimal estimator is $h(Y_1, ..., Y_L) = k_1 Y_1 + k_2 Y_2, ..., k_L Y_L$, for some $k_1, .., k_L \in \mathbb{R}$ if and only if source X and $Z_1, .., Z_L$ are Gaussian. Moreover, if $F_X(\omega)$ and $F_{Z_1}(\omega), .., F_{Z_L}(\omega)$ are constrained to be analytic, then this result also holds for the L_p norm.*

Proof: Since the "if" part is trivial given the previous discussion, we will only prove the "only if" part. We will first focus on two source case, i.e., $L = 2$ and the distortion measure MSE. Let us write the expression for the optimal estimator $\mathbb{E}\{X|Y_1,Y_2\}$ using the Bayes' rule:

$$h(y_1,y_2)=\frac{\int xf_X(x)f_Z(y_1-x,y_2-x)dx}{\int f_X(x)f_Z(y_1-x,y_2-x)dx} \quad (5)$$

Using the fact that Z_1 and Z_2 are independent, we have the following condition for linearity of optimal estimation by setting $h(y_1,y_2)=k_1y_1+k_2y_2$ in (5):

$$[k_1y_1+k_2y_2]\int f_X(x)f_{Z_1}(y_1-x)f_{Z_2}(y_2-x)dx = \int xf_X(x)f_{Z_1}(y_1-x)f_{Z_2}(y_2-x)dx \quad (6)$$

Taking the Fourier transform of both sides and via change of variables $u_1=y_1-x$ and $u_2=y_2-x$ we obtain (7), which can be expressed as

$$k_1\frac{F'_{Z_1}(\omega_1)}{F_{Z_1}(\omega_1)}+k_2\frac{F'_{Z_2}(\omega_2)}{F_{Z_2}(\omega_2)}=(1-k_1-k_2)\frac{F'_X(\omega_1+\omega_2)}{F_X(\omega_1+\omega_2)} \quad (8)$$

Note that the left hand side involves arguments ω_1 and ω_2 while the right hand side argument is $\omega_1+\omega_2$. The only way for this equality to hold for all ω_1,ω_2 and $F_{Z_1}(\omega),F_{Z_2}(\omega),F_X(\omega)$ be valid characteristic functions is that all terms of are linear in terms of their arguments, i.e., $\frac{F'_{Z_1}(\omega_1)}{F_{Z_1}(\omega_1)}=(\log F_{Z_1}(\omega_1))'=\alpha\omega_1,\alpha\in\mathbb{R}$ and so on. This implies that $F_{Z_1}(\omega)=\beta e^{\alpha\omega_1^2}$ for some $\alpha,\beta\in\mathbb{R}$ which is the characteristic function of a Gaussian density. With the same reasoning, it follows that X and Z_2 also have to be Gaussian. It is straightforward to extend the same arguments for $L>2$.

Next, we extend the result to the L_p norm, albeit we now require analyticity of F_X,F_{Z_1} and F_{Z_2}, which means that the moments of X,Z_1 and Z_2 are finite (they have moments of all orders) and moments fully characterize the distribution. The extension to L_p requires a different approach, Lemma 1 plays a key role in the proof. Let us plug three perturbation functions, $\eta(Y_1,Y_2)=Y_1^m,\eta(Y_1,Y_2)=Y_2^m$, and $\eta(Y_1,Y_2)=(Y_1+Y_2)^m$, $m=1,2,..M$ in the necessary condition of Lemma 1, i.e. (4). Then, plugging $Y_i=X+Z_i, i=1,2$ for each of the equations, we obtain a relation between the moments of X,Z_1 and Z_2 with moments of orders up to M. We note that every equation introduces a new variable $\mathbb{E}(X^{m+p-1})$, for $m=1,..,M$, so each new equation is independent of its predecessors. Next, we solve these equations recursively, starting from $m=1$. At each m, we have three unknowns $(\mathbb{E}(X^{m+p-1}),\mathbb{E}(Z_1^{m+p-1}),\mathbb{E}(Z_2^{m+p-1}))$ that are related "linearly". Since the number of linearly independent equations is equal to the number of unknowns for each m, there must exist a unique solution. We know that the moment sequences of the Gaussian source-channel pair satisfy these set of equations since it ensures linearity of optimal estimation. The moment sequence of a Gaussian satisfies Carleman's general criterion [8] and therefore it uniquely determines the corresponding distribution, so the Gaussian source and noise pair is the only solution to these set of equations. ■

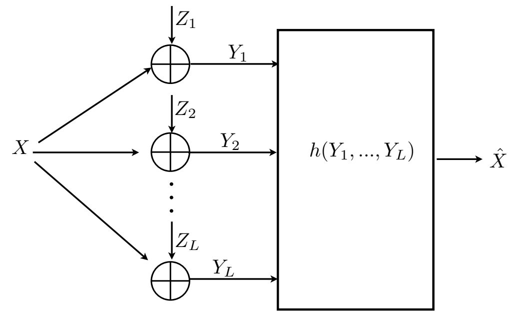

Fig. 2. The "single source multiple measurements" setting

Remark 1. *Note that Theorem 1 is in sharp contrast with the analogous result obtained for the single observation setting, i.e., Result 1. While there are infinitely many source-noise pairs that yield linearity of optimal estimation in the single observation setting, only Gaussian source and noises yield optimality of linear estimation in the multiple observations setting. The surprising nature of the result is captured by the simple example of identically distributed variables $X\sim Z_1\sim Z_2$ where the optimal estimator is not linear despite the fact that $X\sim Z$ in SSSN setting yields linearity of optimal estimation.*

C. Asymptotic Optimality

Theorem 2. *As $\gamma_i\to\infty,\forall i$, the MSE optimal estimator is asymptotically linear if the source X is Gaussian, regardless of the noise distribution. Similarly, in the limit $\gamma_i\to 0,\forall i$, the MSE optimal estimator is asymptotically linear if the noises $Z_1,..,Z_L$ are Gaussian, regardless of the source.*

Proof: We explicitly prove it for $L=2$ case, and the $L>2$ case follows immediately. The proof for $L=2$ applies the central limit theorem [5]. The central limit theorem states that as $\gamma_i\to\infty$, for any finite variance noise Z_i, the functions $F_Z^{\gamma_i}(\omega)$ pointwise converge to the Gaussian characteristic functions. Similarly, as $\gamma_i\to 0$ and for any $F_X(\omega)$, $\frac{1-k_1-k_2}{k_1}\to\infty$ and hence $F_X^{\frac{1-k_1-k_2}{k_1}}(\omega_1+\omega_2)$ converges pointwise to the Gaussian characteristic function and therefore satisfies (8). ■

IV. Multiple Sources-Single Noise Setting

Consider a setting where there are multiple sources, all at the same distance to an interferer, which effectively adds noise to the sources as illustrated in Fig.3. Formally, we want to estimate the independent sources $X_1,X_2,...,X_L$ using the observations $Y_1,..,Y_L$ where $Y_i=X_i+Z$. In this section SNR for i^{th} channel is defined as $\gamma_i=\frac{\sigma^2_{X_i}}{\sigma^2_Z}$.

A. Gaussian Source-Noise Case

If $Z\sim\mathcal{N}(0,\sigma_Z^2)$ and $X_i\sim\mathcal{N}(0,\sigma^2_{X_i})$, the MSE optimal estimator is linear, as in (3) where $k_i=1-\frac{1/\gamma_i}{1+\sum\limits_{j=1}^{L}1/\gamma_j}$. Similar

$$\int [k_1u_1+k_2u_2+(k_1+k_2)x] f_X(x) f_{Z_1}(u_1) f_{Z_2}(u_2) \exp\{-j\omega_1 u_1 - j\omega_2 u_2 - j(\omega_1+\omega_2)x\} dx dy_1 dy_2$$
$$= \int x f_X(x) f_{Z_1}(u_1) f_{Z_2}(u_2) \exp\{-j\omega_1 u_1 - j\omega_2 u_2 - j(\omega_1+\omega_2)x\} dx dy_1 dy_2 \quad (7)$$

to Section IIIA, we have

$$\mathbb{E}\{[X_i - h(Y_1,...,Y_L)]^{p-1}\eta(Y_1,...,Y_L)\} = 0 \quad (9)$$

for any $\eta(Y_1,...,Y_L)$ as a necessary condition for a function $h(Y_1,...,Y_L)$ to be the L_p norm optimal estimator.

B. Main Result

Theorem 3. *The MSE optimal estimator is $h(Y_1,...,Y_L) = k_1Y_1 + k_2Y_2,...,k_LY_L$, for some $k_1,..,k_L \in \mathbb{R}$ if and only if sources $X_1,...,X_L$ and noise Z are Gaussian. Moreover, if $F_{X_1}(\omega),...,F_{X_L}(\omega)$ and $F_Z(\omega)$ are constrained to be analytic, then this result also holds for the L_p norm.*

Proof: Similar to the proof of Theorem 1, we first focus on two source case, i.e., $L = 2$ and MSE. The optimal estimator for X_1 is

$$h_1(y_1,y_2) = \int x f_{X_1|Y_1,Y_2}(x,y_1,y_2)dx \quad (10)$$

Using the fact that X_1 and X_2 are independent and applying the Bayes' rule we obtain

$$h_1(y_1,y_2) = \frac{\int (y_1-z) f_Z(z) f_{X_1}(y_1-z) f_{X_2}(y_2-z)dz}{\int f_Z(z) f_{X_1}(y_1-z) f_{X_2}(y_2-z)dz}$$

Following the same steps as in the proof Theorem 1, we have

$$(\log F_{X_1}^{1-k_1}(\omega_1))' = (\log F_Z(\omega_1+\omega_2))' + (\log F_{X_2}^{k_2}(\omega_2))' \quad (11)$$

We observe that (11) is satisfied only if F_{X_1}, F_{X_2}, F_Z are the characteristic functions of Gaussians, hence X_1, X_2, Z must be Gaussian to ensure linearity of MSE optimal estimator. Extension to the L_p norm follows from applying arguments used in the L_p extension in the proof of Theorem 1, on (9). ■

Remark 2. *An alternative proof of Theorem 3 can be derived using the duality between the SSMM and the MSSN problems. The linear estimator in the MSSN setting can be viewed as: first estimate Z as $\hat{Z} = \sum k_iY_i$ and then, simply estimate X_i from observations $\hat{X}_i = Y_i - \hat{Z}$. This linear estimator is optimal only if both of these estimation steps are optimal. Estimating Z in the MSSN setting is identical to the the SSMM problem with Z as the source and X_i as the noise variables. We know from Theorem 1 that the first step is optimal only if X_i and Z are Gaussian, hence linear estimator is optimal in the MSSN setting only if all variables are Gaussian.*

Theorem 4. *In the limit $\gamma_i \to 0, \forall i$, the MSE optimal estimator is asymptotically linear if the noise is Gaussian, regardless of the source distributions. Similarly, as $\gamma_i \to \infty, \forall i$, the MSE optimal estimator is asymptotically linear if the sources are Gaussian, regardless of the noise.*

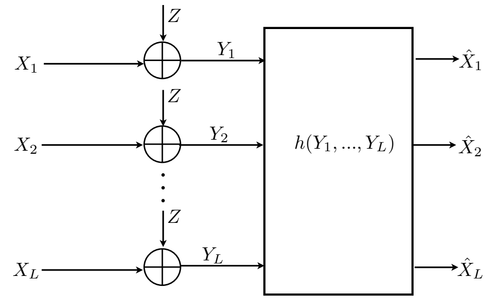

Fig. 3. The "multiple sources, single noise" setting

The proof is left out for brevity.

V. Conclusion

In this paper, we considered two common estimation settings: i) source estimation from multiple noisy measurements, and ii) estimating multiple independent sources, all of which are corrupted by the same noise realization. For both of these settings, we showed that the optimal estimator is linear if and only if sources and noises are Gaussian. This is in sharp contrast with the single source-single measurement or the associated vector settings, where there exist infinitely many source-noise pairs that guarantee linearity of optimal estimation. We also showed, for both settings, that at asymptotically high SNR, the optimal estimator for Gaussian source(s) converges to linear, irrespective of the noise distribution(s); and at asymptotically low SNR, the optimal estimator is linear for Gaussian noise(s), regardless of the source(s).

References

[1] C. Rothschild and E. Mourier, "Sur les lois de probabilité à regression linéaire et écart type lié 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] 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.

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

[6] E. Akyol, K. Viswanatha, and K. Rose, "On conditions for linearity of optimal estimation," *IEEE Transactions on Information Theory, Jul 2012*.

[7] E. Akyol, K. Viswanatha, and K. Rose, "On multidimensional optimal estimators: Linearity conditions," in *IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2011, pp. 741–744.

[8] J.A. Shohat and J.D. Tamarkin, *The Problem of Moments*, American Mathematical Society, 1943.