On Regret of Parametric Mismatch in Minimum Mean Square Error Estimation

Majid Fozunbal

HP Laboratories
HPL-2010-10

Abstract:
This paper studies the effect of parametric mismatch in minimum mean square error (MMSE) estimation. In particular, we consider the problem of estimating the input signal from the output of an additive white Gaussian channel whose gain is fixed, but unknown. The input distribution is known, and the estimation process consists of two algorithms. First, a channel estimator blindly estimates the channel gain using past observations. Second, a mismatched MMSE estimator, optimized for the estimated channel gain, estimates the input signal. We analyze the regret, i.e., the additional mean square error, that is raised in this process. We derive upper-bounds on both absolute and relative regrets. Bounds are expressed in terms of the Fisher information. We also study regret for unbiased, efficient channel estimators, and derive a simple trade-off between Fisher information and relative regret. This trade-off shows that the product of a certain function of relative regret and Fisher information equals the signal-to-noise ratio, independent of the input distribution. The trade-off relation implies that higher Fisher information results to smaller expected relative regret.

On Regret of Parametric Mismatch in Minimum Mean Square Error Estimation

Majid Fozunbal

Hewlett-Packard Laboratories

Palo Alto, CA 94304

Abstract— This paper studies the effect of parametric mismatch in minimum mean square error (MMSE) estimation. In particular, we consider the problem of estimating the input signal from the output of an additive white Gaussian channel whose gain is fixed, but unknown. The input distribution is known, and the estimation process consists of two algorithms. First, a channel estimator blindly estimates the channel gain using past observations. Second, a mismatched MMSE estimator, optimized for the estimated channel gain, estimates the input signal. We analyze the *regret*, i.e., the additional mean square error, that is raised in this process. We derive upper-bounds on both absolute and relative regrets. Bounds are expressed in terms of the Fisher information. We also study regret for unbiased, efficient channel estimators, and derive a simple trade-off between Fisher information and relative regret. This trade-off shows that the product of a certain function of relative regret and Fisher information equals the signal-to-noise ratio, independent of the input distribution. The trade-off relation implies that higher Fisher information results to smaller expected relative regret.

I. Introduction

Consider an application that you are given the output of a system, and you seek to recover the input of the system. You know that the system is noisy, e.g., it adds white Gaussian noise to the output. You know the distribution of the input, but you do not know the system parameters. Problems of this sort arise in different applications in signal processing and communication systems. Some examples include blind deconvolution [1], dereverberation [2], denoising [3], and mismatch decoding [4]. These applications differ in their fundamental models, fidelity criteria, and methodologies. However, they have one thing in common: they all suffer from parametric mismatch in recovering the input signals.

The motivation of this work is blind deconvolution and dereverberation applications. Linear time-invariant channels serve as common models in these applications. As the input signal passes through these channels, it convolves with the unknown finite-impulse response (FIR) of the channel, and it adds with additive white Gaussian noise (of known variance). Recovering the input signals from the noisy output could be impossible even with perfect knowledge about the channel response. This is out of our scope. Instead, we aim to study the penalty and performance degradation that is specifically caused by the lack of knowledge about the channel response.

We benchmark performance against that of perfect channel knowledge scenario. We are concerned about issues such as required sample complexity or training in channel estimation

to bring performance of input estimation within a desired range. As a counterpart problem in communication systems, one may think of block fading channels and the trade-off between accuracy of channel estimation and performance of decoding [5]. Note that channel estimation in our case is blind as we have no control of the source.

As a first step to address these problems, in this work, we focus on the most basic system in which the unknown channel is just a single gain. We expect that the results and intuitions of this work will generalize to generic FIR channels.[1] In treating the problem, we consider an estimation process that consists of two algorithms. First, a *channel estimator* blindly estimates the channel gain using past output observations. Second, a mismatched *minimum mean square error* (MMSE) estimator, optimized for the estimated channel gain, estimates the input signal. Figure 1 illustrates the building blocks of this process. Due to estimation error in channel estimation, the MMSE estimator that is used in recovering the input signal results in a mean square error that is larger than that of the ideal MMSE estimator. We call this additional error as *regret*, and we derive novel upper-bounds on both absolute and relative regrets. The bounds are simple and demonstrate interesting connections to the Fisher information. To this end, one might attempt to exploit the results of [6] and [7] to derive other alternative bounds.

We also quantify regret for unbiased, *efficient* channel estimators. Since these estimators achieve Cramer-Rao bound, they result in a simple trade-off relation between Fisher information and relative regret. This trade-off relation expresses that the product of a certain function of relative regret and Fisher information is equivalent to the signal-to-noise ratio, independent of the input distribution. Trade-off suggests that higher Fisher information results to smaller expected relative regret. Although, intuitively, this may seem expected, simplicity of the trade-off relation makes it worthwhile.

II. Setup

Consider a linear dynamic system

$$Y_n = aX_n + V_n \tag{1}$$

in which $\{V_n\}$ is an independent, identically, distributed (i.i.d.) Gaussian noise such that $V_n \sim \mathcal{N}(0, \sigma_v^2)$. The input X_n is

---

[1] Analogous to the case between the analysis of flat-fading and the analysis of frequency-selective channels.

an i.i.d. process whose distribution is known to be $P(X)$. Parameter $a \in \mathbb{R}^+$ is a fixed, unknown channel gain. It results to a derived parametric family of probability measures $P_a(X, Y)$, the joint distribution of X and Y, governing the system dynamic (1). The objective is to observe a realization of the output process

$$Y^n = (Y_1, Y_2, \cdots, Y_n)$$

and estimate the realization of the underlying input process, i.e.,

$$X^n = (X_1, X_2, \cdots, X_n).$$

Let $\mathcal{X} = \mathbb{R}$ and $\mathcal{Y} = \mathbb{R}$ denote the input and output spaces, respectively. We consider memoryless input estimators, e.g., $\phi \colon \mathcal{Y} \to \mathcal{X}$ where $\phi(Y_n)$ is an estimate for X_n. The *mean square error* (MSE) for ϕ is defined

$$\mathbb{E}\left[(X - \phi(Y))^2\right] = \int (x - \phi(y))^2 dP_a. \tag{2}$$

In Eq. (2) and henceforth we follow the convention that unsubscribed expectations are measured according to $P_a(X, Y)$. Moreover, we use concise notations like $P_a = P_a(X, Y)$ and $P_{a|y} = P_a(X|Y = y)$ to denote joint and conditional distributions, respectively.

One seeks to find an estimator that minimizes MSE (2). The main challenge, however, is that a and P_a are unknown. Should we had oracle knowledge about a, the MMSE estimator for X is defined

$$\phi_a(y) = \mathbb{E}[X|Y = y]. \tag{3}$$

for an observation $Y = y$. Any other estimator ϕ results to additional error that we call it *regret*. The motivation for this name is that it measures degradation on performance, an impact caused by imprecise knowledge about a.

In this paper, we assess regret for a special class of mismatched estimators. Namely, we consider an estimation process that is depicted in Figure 1. A channel estimation works in parallel with an MMSE input estimation as follows. At time instance n, a channel estimator finds an estimate $\hat{a} = \hat{a}_n$ of a using the observed values Y^{n-1}. Then, it uses the optimal estimator of $P_{\hat{a}}(X, Y)$ to compute

$$\phi_{\hat{a}}(y_n) = \mathbb{E}_{\hat{a}}[X_n|Y = y_n] \tag{4}$$

as an estimate for X_n. Function $\phi_{\hat{a}}$ is a mismatch MMSE estimator that causes regret when used in place of ϕ_a. In the following sections, we study two types of regret: *absolute regret* and *relative regret*.

III. ABSOLUTE REGRET

A. Deviation Analysis

The absolute regret corresponding to $\phi_{\hat{a}}$ is

$$R(\hat{a}, a) = \mathbb{E}\left[(X - \phi_{\hat{a}}(Y))^2\right] - \mathbb{E}\left[(X - \phi_a(Y))^2\right]. \tag{5}$$

Application of orthogonality principle results to

$$R(\hat{a}, a) = \mathbb{E}\left[(\phi_{\hat{a}}(Y) - \phi_a(Y))^2\right]. \tag{6}$$
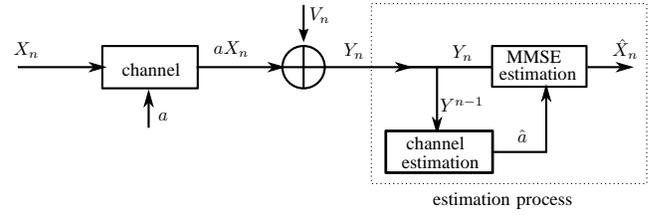
Fig. 1. Figure depicts the building blocks of the system setup. The estimation process consists of two individual algorithms: 1) a channel estimation algorithm that blindly estimates $\hat{a}$ as an estimate for a, 2) a mismatch MMSE estimation algorithm, optimized for $\hat{a}$, that recovers X_n.

Eq. (6) quantifies the absolute *regret* of using $\phi_{\hat{a}}$ instead of ϕ_a. The following lemma states and proves an upper-bound on (6).

Lemma 3.1: For every $\hat{a}$, the following holds true

$$R(\hat{a}, a) \le (\hat{a} - a)^2 \mathbb{E}\left[k(Y) J(X; a|Y = y)\right]$$
$$+ o(\hat{a} - a)^2 \tag{7}$$

in which the expectation is with respect to Y,

$$k(Y) \triangleq 6\sigma_x^2 + 4\frac{Y^2}{a^2} + 4\frac{Y^2}{\hat{a}^2},$$

and

$$J(X; a|Y = y) \triangleq \mathbb{E}\left[\left(\nabla \ln f_a(X|Y = y)\right)^2\right] \tag{8}$$

is the Fisher information of X relative to a, conditioned on $Y = y$. Here, $f_a(X|Y = y)$ is the density of $P_{a|y}$.

Proof: Refer to Appendix A. ∎

Lemma 3.1 describes a bound (7) that comprises two multiplicative terms. The first term $(\hat{a} - a)^2$ measures the channel estimation error. The second term is the weighted average of conditional Fisher information. Intuitively, this term measures the amount of information that an observable random variable X carries about unknown parameter a conditioned on Y, assigning more weight to larger values of Y.

Corollary 3.1: For $|\hat{a} - a| << 1$, we have

$$R(\hat{a}, a) \le (\hat{a} - a)^2 \mathbb{E}[k(Y) J(X; a|Y = y)]. \tag{9}$$

Moreover, if $J(X; a|Y = y)$ and $k(Y)$ are uncorrelated, we obtain the simple bound

$$R(\hat{a}, a) \le (\hat{a} - a)^2 (14\sigma_x^2 + 8\frac{\sigma_v^2}{a^2}) J(X; a|Y) \tag{10}$$

in which $J(X; a|Y) \triangleq \mathbb{E}[J(X; a|Y = y)]$.

B. Efficient Channel Estimation

Note that (9) does not depend on the channel estimation algorithm that estimates a. It simply relates small deviation between $\hat{a}$ and a to absolute regret in estimating X. To incorporate the effect of channel estimation algorithm, we proceed as follows.

As mentioned earlier, at time n, $\hat{a}$ is obtained through observation of $Y^{n-1} = (Y_i)_{i=1}^{n-1}$. In formal terms,

$$\hat{a} = A_n(Y^{n-1})$$

where $A = (A_1, A_2, \cdots)$ is a channel estimation algorithm in which $A_n \colon \mathcal{Y}^{n-1} \to \mathbb{R}^+$.

Lemma 3.2: Let $\mathcal{A}$ denote the class of all unbiased channel estimation algorithms. If $\mathcal{A}$ contains an *efficient estimator* [8, p. 92], the following holds true

$$\inf_{A \in \mathcal{A}} \mathbb{E}\left[R(A_n(Y^{n-1}), a)\right] \leq$$
$$\frac{1}{n-1} \frac{\mathbb{E}\left[k(Y)J(X;a|Y)\right]}{J(Y;a)} \quad (11)$$

for sufficiently large values of n.[2]

Proof: Refer to Appendix B. ∎

IV. RELATIVE REGRET

A. Deviation Analysis

Let

$$RR(\hat{a}, a) = \mathbb{E}\left[\frac{(\phi_{\hat{a}}(Y) - \phi_a(Y))^2}{\mathbb{E}_{\hat{a}}\left[X^2|Y\right] + \mathbb{E}_a\left[X^2|Y\right]}\right] \quad (12)$$

denote the relative regret. The following lemma states and proves a simple upper-bound on $RR(\hat{a}, a)$.

Lemma 4.1: For every $\hat{a}$, we have

$$RR(\hat{a}, a) \leq (\hat{a} - a)^2 \; J(X;a|Y) + o(\hat{a} - a)^2 \quad (13)$$

where

$$J(X;a|Y) \triangleq \mathbb{E}\left[\left(\nabla \ln f_a(X|Y)\right)^2\right] \quad (14)$$

is the Fisher information of X relative to a, conditioned on Y.

Proof: See Appendix C. ∎

The novelty of Eq. (13) is that it expresses a simple upper-bound on the relative regret for small deviations between $\hat{a}$ and a. For small $(\hat{a} - a)^2$, it simplifies to

$$RR(\hat{a}, a) \leq (\hat{a} - a)^2 \; J(X;a|Y). \quad (15)$$

B. Efficient Channel Estimation

Similar to the case for absolute regret, we now state the following result.

Lemma 4.2: Let $\mathcal{A}$ denote the class of all unbiased estimation algorithms. If $\mathcal{A}$ contains an *efficient estimator*, the following holds true

$$\inf_{A \in \mathcal{A}} \mathbb{E}\left[RR(A_n(Y^{n-1}), a)\right] \leq \frac{1}{n-1} \frac{J(X;a|Y)}{J(Y;a)} \quad (16)$$

for sufficiently large values of n.

Proof: The proof of this lemma is essentially the same as the proof of Lemma 3.2. ∎

Lemma 4.2 describes a bound on the expected relative regret, should an efficient estimator be used. This bound determines the smallest upper-bound on average relative regret, when sufficiently good unbiased channel estimators are used.

[2]The expectation in the LHS is with respect to Y^{n-1}.

C. Regret Scalar

The constant value in the RHS of Eq. (16) worths attention. It does not change with respect to n, and as $n \to \infty$, it becomes the sole scalar that determines the level of relative regret. We define this quantity as the *regret scalar* and denote it by

$$\rho(a) = \frac{J(X;a|Y)}{J(Y;a)}. \quad (17)$$

Lemma 4.3: For every zero-mean input distribution $P(X)$, the following trade-off holds true between regret scalar and output fisher information

$$(\rho(a) + 1)J(Y;a) = \frac{\sigma_x^2}{\sigma_v^2}. \quad (18)$$

Proof: See Appendix D. ∎

In Eq. (18), the RHS is the signal-to-noise ratio that is independent of a. Thus, Eq. (18) presents a simple product trade-off relationship between $\rho(a)$ and $J(Y;a)$. It suggest that the higher the fisher information, the smaller the regret scalar, and vice-versa. The following example explicates this trade-off.

Example 4.1 (Gaussian Input): Assume $X_n \sim \mathcal{N}(0, \sigma_x^2)$ and $V_n \sim \mathcal{N}(0, \sigma_v^2)$ are i.i.d. implying that $Y_n \sim \mathcal{N}(0, a^2\sigma_x^2 + \sigma_v^2)$ and $Y_n|x_n \sim \mathcal{N}(ax_n, \sigma_v^2)$. With perfect knowledge of a, the ideal estimator for X given $Y = y$ is

$$\phi_a(y) = \frac{a\sigma_x^2}{a^2\sigma_x^2 + \sigma_v^2}y. \quad (19)$$

The MMSE error resulting from this estimator is

$$\mathbb{E}\left[(X - \phi_a(Y))^2\right] = \frac{\sigma_x^2\sigma_v^2}{a^2\sigma_x^2 + \sigma_v^2}. \quad (20)$$

A mismatch estimator for $\hat{a}$ is

$$\phi_{\hat{a}}(y) = \frac{\hat{a}\sigma_x^2}{\hat{a}^2\sigma_x^2 + \sigma_v^2}y. \quad (21)$$

We have

$$J(Y;a|X) = \frac{\sigma_x^2}{\sigma_v^2} \quad (22)$$

and

$$J(Y;a) = \frac{2a^2\sigma_x^4}{(a^2\sigma_x^2 + \sigma_v^2)^2}. \quad (23)$$

Thus,

$$\rho(a) = \frac{1}{2}\left(\frac{a^2\sigma_x^2}{\sigma_v^2} + \frac{\sigma_v^2}{a^2\sigma_x^2}\right) \quad (24)$$

Figure 2 depicts the behavior of $\rho(a)$ and $J(Y;a)$ with respect to a. The SNR $= \frac{\sigma_x^2}{\sigma_v^2} = 10$ dB and at $a = .35$, the minimum regret scalar coincides with maximum Fisher information.

V. RECAP AND CONCLUSION

We considered the problem of estimating the input signal from the output of an additive white Gaussian noise channel

Fig. 2. Figure illustrates the multiplicative trade-off between regret scalar and Fisher information. Smaller Fisher information results to larger regret scalar and vice versa. The $\mathrm{SNR} = 10$ dB and the minimum regret scalar is coincident with maximum Fisher information.

subject to parametric uncertainty. Namely, the channel gain is fixed, but unknown. In treating the problem, we considered an estimation process that consists of two algorithms: a blind channel estimator and a mismatched MMSE estimator to estimate the input. We studied the regret that is raised as a result of mismatch estimation. Simple upper-bounds on both absolute and relative regrets were presented. These bounds provide useful tools in assessing deviation in estimating the input when there exists a small deviation in channel gain estimation. The bounds are simple and expressed in terms of the Fisher information. This makes them more intuitive and could potentially bridge to other known results in the literature.

We also quantified regret for unbiased, efficient channel estimators. Using Caramer-Rao bound, we derived a simple trade-off between Fisher information and relative regret. This trade-off expresses that the product of a certain function of relative regret and the fisher information is equivalent to the signal-to-noise ratio, independent of the input distribution. The trade-off suggests that the higher the Fisher information, the smaller the expected relative regret.

This work is our initial attempt to shed light on information-theoretic limits of blind deconvolution and dereverberation systems. We are currently working on generalization of these results to these applications.

APPENDIX

A. Proof of Lemma 3.1

To derive an upperbound on absolute regret, we first state and prove the following results.

Proposition 1.1: For every $\hat{a}$ and $y \in \mathcal{Y}$, we have

$$
(\phi_{\hat{a}}(y) - \phi_a(y))^2 \leq
$$
$$
2(\mathbb{E}_{\hat{a}}\left[X^2|y\right] + \mathbb{E}_a\left[X^2|y\right])D(P_{\hat{a}|y}\|P_{a|y}). \quad (25)
$$

Proof: By definition, we have

$$
(\phi_{\hat{a}}(y) - \phi_a(y))^2 = \left(\int x\big(\frac{dP_{\hat{a}|y}}{dQ} - \frac{dP_{a|y}}{dQ}\big)dQ\right)^2
$$

for every probability measure Q such that $P_{a|y} \ll Q$ and $P_{\hat{a}|y} \ll Q$. By Cauchy Schwartz inequality, we obtain

$$
(\phi_{\hat{a}}(y) - \phi_a(y))^2 \leq \int x^2 \left(\sqrt{\frac{dP_{\hat{a}|y}}{dQ}} + \sqrt{\frac{dP_{a|y}}{dQ}}\right)^2 dQ
$$
$$
\cdot \int \left(\sqrt{\frac{dP_{\hat{a}|y}}{dQ}} - \sqrt{\frac{dP_{a|y}}{dQ}}\right)^2 dQ \quad (26)
$$

By inequality $(a + b)^2 \leq 2(a^2 + b^2)$, one can show that the first term in the RHS of the above inequality is smaller than or equal to

$$
2(\mathbb{E}_a\left[X^2|y\right] + \mathbb{E}_{\hat{a}}\left[X^2|y\right]).
$$

The second term in the RHS of inequality (26) is known as *Kakutani-Hellinger* distance between $P_a(X|y)$ and $P_{\hat{a}}(X|y)$, denoted by [9, p. 363]

$$
r^2(P_{\hat{a}|y}, P_{a|y}) = \frac{1}{2} \int \left(\sqrt{\frac{dP_{\hat{a}|y}}{dQ}} - \sqrt{\frac{dP_{a|y}}{dQ}}\right)^2 dQ.
$$

Moreover, we know of the following inequality between Kakutani-Hellinger distance and Kullback-Leibler distance [9, p. 369]

$$
2r^2(P_{\hat{a}|y}, P_{a|y}) \leq D(P_{\hat{a}|y}\|P_{a|y}).
$$

Substituting in (26), we obtain Eq. (25). ∎

Proposition 1.2: For every $\hat{a}$ and $y \in \mathcal{Y}$, the following inequality holds true

$$
\mathbb{E}_{\hat{a}}\left[X^2|y\right] \leq 3\sigma_x^2 + 4\frac{y^2}{\hat{a}^2}. \quad (27)
$$

Proof: Let $f_{\hat{a}}(y|x)$ and $f_{\hat{a}}(y)$ denote the conditional and marginal densities for $P_{\hat{a}}(X,Y)$. Then,

$$
\mathbb{E}_{\hat{a}}\left[X^2|y\right] = \int x^2 \frac{f_{\hat{a}}(y|x)}{f_{\hat{a}}(y)}f(x)dx
$$
$$
= \int_{x:f_{\hat{a}}(y|x)\leq f_{\hat{a}}(y)} + \int_{x:f_{\hat{a}}(y|x)> f_{\hat{a}}(y)}
$$
$$
\leq \mathbb{E}\left[X^2\right] + \int_{x:f_{\hat{a}}(y|x)> f_{\hat{a}}(y)} \quad (28)
$$

To simplify the second term, we substitute x^2 by the inequality that is derived as follows

$$
f_{\hat{a}}(y|x) > f_{\hat{a}}(y) \Rightarrow
$$
$$
(y - \hat{a}x)^2 < -2\sigma_v^2 \ln\left(\sqrt{2\pi}\sigma_v f_{\hat{a}}(y)\right).
$$

Taking the square roots, we obtain

$$|y - \hat{a}x| < \sqrt{-2\sigma_v^2 \ln\left(\sqrt{2\pi}\sigma_v f_{\hat{a}}(y)\right)} \Rightarrow$$

$$|\hat{a}x| < |y| + \sqrt{-2\sigma_v^2 \ln\left(\sqrt{2\pi}\sigma_v f_{\hat{a}}(y)\right)}.$$

Taking the square of both sides of the previous inequality and using the inequality $(a + b)^2 \leq 2(a^2 + b^2)$, we obtain

$$x^2 < 2\frac{y^2}{\hat{a}^2} - 4\frac{\sigma_v^2}{\hat{a}^2}\ln\left(\sqrt{2\pi}\sigma_v f_{\hat{a}}(y)\right) \Rightarrow$$

$$x^2 < 2\frac{y^2}{\hat{a}^2} + 4\frac{\sigma_v^2}{\hat{a}^2}\int \frac{(y - \hat{a}x)^2}{2\sigma_v^2}f(x)dx \Rightarrow$$

$$x^2 < 4\frac{y^2}{\hat{a}^2} + 2\sigma_x^2$$

By substituting for x^2 in the second term of the RHS of Eq. (28), we conclude Eq. (27). ∎

As a result of Propositions 1.1 and 1.2, we obtain

$$(\phi_{\hat{a}}(y) - \phi_a(y))^2 \leq 2(6\sigma_x^2 + 4\frac{y^2}{\hat{a}^2} + 4\frac{y^2}{a^2})D(P_{\hat{a}|y}\|P_{a|y}).$$

Moreover, the following equality is known between Kullback-Leibler distance and Fisher information [10, p.55]

$$D(P_{\hat{a}|y}\|P_{a|y}) = \frac{(\hat{a} - a)^2}{2}J(X; a|Y = y) + o(\hat{a} - a)^2, \qquad (29)$$

where

$$J(X; a|Y = y) \triangleq \mathbb{E}\left[\left(\nabla \ln f_a(X|Y = y)\right)^2\right]$$

is the Fisher information of X relative to a, conditioned on $Y = y$. Defining

$$k(y) \triangleq 6\sigma_x^2 + 8\frac{y^2}{a^2}$$

and taking expectation with respect to Y, we conclude the proof of Lemma 3.1.

B. Proof of Lemma 3.2

We know that $\hat{a} = A_n(Y^{n-1})$. For sufficiently large values of n, Eq. (9) holds true with arbitrarily high probability. Taking the expectation of both sides of Eq. (9) with respect to Y^{n-1}, we obtain

$$\mathbb{E}\left[R(A_n(Y^{n-1}), a)\right] \leq$$
$$\mathbb{E}\left[(A_n(Y^{n-1}) - a)^2\right]\mathbb{E}\left[k(Y)J(X; a|Y)\right]$$

Take the infimum of both sides over $\mathcal{A}$ and assume $\mathcal{A}$ contains an *efficient estimator* [8, p. 92]. By definition an efficient estimator achieves the Cramer-Rao bound. This means

$$\mathbb{E}\left[(A_n(Y^{n-1}) - a)^2\right] = \frac{1}{J(Y^{n-1}; a)}.$$

Since Y_n is i.i.d., by additivity of Fisher information

$$J(Y^{n-1}; a) = (n - 1)J(Y; a).$$

As a result, we obtain

$$\inf_{A \in \mathcal{A}} \mathbb{E}\left[R(A_n(Y^{n-1}), a)\right] \leq$$
$$\frac{1}{n - 1}\frac{\mathbb{E}\left[k(Y)J(X; a|Y)\right]}{J(Y; a)}.$$

C. Proof of Lemma 4.1

By Proposition 1.1, we have

$$\frac{(\phi_{\hat{a}}(y) - \phi_a(y))^2}{\mathbb{E}_{\hat{a}}\left[X^2|y\right] + \mathbb{E}_a\left[X^2|y\right]} \leq 2D(P_{\hat{a}|y}\|P_{a|y}).$$

By Eq. (29), we obtain

$$RR(\hat{a}, a) \leq (\hat{a} - a)^2\mathbb{E}\left[\left(\nabla \ln f_a(X|Y)\right)^2\right] + o(\hat{a} - a)^2,$$

where

$$J(X; a|Y) \triangleq \mathbb{E}\left[\left(\nabla \ln f_a(X|Y)\right)^2\right]$$

is the average Fisher information of X relative to a, conditioned on Y.

D. Proof of Lemma 4.3

Since X does not depend on a, $J(X; a) = 0$, and hence

$$\rho(a) = \frac{J(X; a|Y)}{J(Y; a)} = \frac{J(Y; a|X)}{J(Y; a)} - 1.$$

Moreover, since the additive noise is Gaussian, the equality

$$J(Y; a|X) = \frac{\sigma_x^2}{\sigma_v^2}$$

holds true for every distribution $P(X)$ with zero mean. As a result, we obtain Eq. (18).

REFERENCES

[1] D. Donoho, "On minimum entropy deconvolution," *in Applied time series analysis*, pp. 565–608, 1981, editor D. Finley.

[2] M. Fozunbal, "Performance analysis for blind identification of acoustic channels," *Proc. IEEE Intrn. Conf. Acoust, Speech, and Sig. Proc. (ICASSP)*, Mar. 2010, to appear.

[3] G. Gemelos, S. Sigurjonsson, and T. Weissman, "Universal minimax discrete denoising under channel uncertainty," *IEEE Trans. Inform. Theory*, vol. 52, no. 8, pp. 3476–3497, Aug. 2006.

[4] A. Ganti, A. Lapidoth, , and E. Telatar, "Mismatched decoding revisited: General alphabets, channels with memory, and the wide-band limit," *IEEE Trans. Inform. Theory*, vol. 46, no. 7, pp. 2315–2328, Nov. 2000.

[5] B. Hassibi and B. M. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inform. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.

[6] S. Verdu, "Mismatched estimation and relative entropy," *Proc. IEEE Intrn. Symp. Inform. Theory (ISIT)*, Jun. 2009.

[7] D. Guo, "Relative entropy and score function: New informationestimation relationships through arbitrary additive perturbation," *Proc. IEEE Intrn. Symp. Inform. Theory (ISIT)*, Jun. 2009.

[8] A. A. Borovkov, *Mathematical Statistics*, A. Moullagaliev, Ed. Gordon and Breach Science Publishers, 1998.

[9] A. N. Shiryaev, *Probability*, 2nd ed., ser. Graduate Texts in Mathematics, R. P. Boas, Ed. New York: Springer-Verlag, 1996, no. 95.

[10] S. Kullback, *Information Theory and Statistics*. New York: Dover, 1968.