Conference Paper

Analysis of Mismatched Estimation Errors Using Gradients of Partition Functions

Author(s):
Huleihel, Wasim; Merhav, Neri

Publication Date:
2014

Permanent Link:
https://doi.org/10.3929/ethz-a-010094593

Rights / License:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
Analysis of Mismatched Estimation Errors Using Gradients of Partition Functions

Wasim Huleihel and Neri Merhav
Technion - Israel Institute of Technology
Department of Electrical Engineering
Haifa 32000, ISRAEL
E-mail: {wh@tx, merhav@ee}.technion.ac.il

Abstract—We consider the problem of signal estimation (de-noising) from a statistical-mechanical perspective, in continuation to a recent work on the analysis of mean-square error (MSE) estimation using a direct relationship between optimum estimation and certain partition functions. Accordingly, we derive a single-letter expressions of the MMSE and mismatched MSE of a codeword (from a randomly selected code), corrupted by a Gaussian vector channel, and we provide several examples to demonstrate phase transitions in the behavior of the MSE.

I. INTRODUCTION

The connections and the interplay between information theory, statistical physics and signal estimation have been known for several decades, and they are still being studied from a variety of aspects, see, for example [1-3] and many references therein.

Recently, in [2], the well known I-MMSE relation [3], which relates the mutual information and the derivative of the minimum mean-square error (MMSE), was further explored using a statistical physics perspective. One of the main contributions in [2] is the demonstration of the usefulness of statistical-mechanical tools (in particular, utilizing the fact that the mutual information can be viewed as the partition function of a certain physical system) in assessing MMSE via the I-MMSE relation of [3]. More recently, Merhav [1] proposed a more flexible method, whose main idea is that, for the purpose of evaluating the covariance matrix of the MMSE estimator, one may use other information measures, which have the form of a partition function and hence can be analyzed using methods of statistical physics (see, e.g., [4] and many references therein). The main advantage of the proposed approach over the I-MMSE relations, is its full generality: Any joint probability function \( P(x, y) \), where \( x \) and \( y \) designate the channel input to be estimated and the channel output, respectively, can be handled (for example, the channel does not have to be additive or Gaussian). Moreover, using this approach, any mismatch, both in the source and the channel, can be considered.

This paper is a further development of [1] in the above described direction. Particularly, in [1, Section IV. A], the problem of mismatched estimation of a codeword, transmitted over an additive white Gaussian (AWGN) channel, was considered. It was shown that the mismatched MSE exhibits phase transitions at some rate thresholds, which depend upon the real and the mismatched parameters of the problem, and the behavior of the receiver. To wit, the mismatched MSE acts inherently differently for a pessimistic and optimistic receivers, where in the example considered in [1, Section IV. A] pessimism literally means that the estimator assumes that the channel is worse than it really is (in terms of signal-to-noise ratio (SNR)), and the vice versa for optimism. In this paper, we extend the above described model to a much more general one; the Gaussian vector channel, which has a plenty of applications in communications and signal processing. It is important to emphasize that compared to [1, 2], the mathematical analysis is much more complicated (consisting of some new concepts), and the notions of pessimism and optimism described above, also play a significant role in this model, although their physical meanings in general are not obvious. Moreover, in contrast to previous work on mismatched estimation, the case of channel mismatch is explored, namely, the receiver has a wrong assumption on the channel.

II. MODEL AND PROBLEM FORMULATION

Let \( C = \{x_0, \ldots, x_{M-1}\} \) denote a codebook of size \( M = e^{nR} \), which is selected at random (and then revealed to the estimator) in the following manner: Each \( x_i \) is drawn independently under the uniform distribution over the surface of the \( n \)-dimensional hypersphere, which is centered at the origin, and whose radius is \( \sqrt{nR} \). Finally, let \( X \) assume a uniform distribution over \( C \). We consider the Gaussian vector channel model

\[
Y = AX + N,
\]

where \( Y, X \) and \( N \) are random vectors in \( \mathbb{R}^n \), designating the channel output vector, the transmitted codeword and the noise vector, respectively. It is assumed that the components of the noise vector, \( N_i \), are i.i.d., zero-mean, Gaussian random variables with variance \( 1/\beta \), where \( \beta \) is a given positive constant designating the signal-to-noise ratio (SNR) (for \( P_s = 1 \)), or the inverse temperature in the statistical-mechanical jargon. We further assume that \( X \) and \( N \) are statistically independent. Finally, the channel matrix, \( A \in \mathbb{R}^{n \times n} \), is assumed to be a given deterministic Toeplitz matrix, whose entries are given by the coefficients of the impulse response of a given linear system. Specifically, let \( \{h_k\} \) denote the generating sequence (or impulse response) of \( A \), so that \( A = \{a_{i,j}\}_{i,j} = \{h_{i-j}\}_{i,j} \), and let \( H(\omega) \) designate the frequency response (Fourier transform) of \( \{h_k\} \).

There are several motivations for codeword estimation. One example is that of a user that, in addition to its desired signal, receives also a relatively strong interference signal, which
carries digital information intended to other users, and which comes from a codebook whose rate exceeds the capacity of this crosstalk channel between the interferer and the user, so that the user cannot fully decode this interference. Nevertheless, our user would like to estimate the interference as accurately as possible for the purpose of cancellation. Furthermore, we believe that the tools/concepts developed in this paper for handling matched and mismatched problems, can be used in other applications in signal processing and communication. Such examples are denoising, mismatched decoding, blind deconvolution, and many other applications. Note that although the aforementioned examples are radically different (in terms of their basic models and systematization), they will all suffer from mismatch when estimating the input signals.

As was mentioned previously, we analyze the problem of mismatched codeword estimation which is formulated as follows: Consider a mismatched estimator which is the conditional mean of \( X \) given \( Y \), based on an incorrect joint distribution \( P' (x, y) \), whereas the true joint distribution continues to be \( P (x, y) \). Accordingly, the mismatched MSE is defined as

\[
\text{mse} (X | Y) = \frac{1}{T} E \| X - E' (X | Y) \|^2
\]

where \( \text{mse} (X | Y) \) is the conditional expectation with respect to (w.r.t.) the mismatched measure \( P' \). In this paper, the following mismatch mechanism is assumed: The input measure is matched, i.e., \( P (x) = P' (x) \) (namely, the mismatched estimator knows the true code), both conditional distributions (“channels”) \( P' (\cdot | x) \) and \( P' (\cdot | x) \) are Gaussian, but are associated with different channel matrices. More precisely, while the true channel matrix (under \( P \)) is \( A \), the assumed channel matrix (under \( P' \)) is \( A' \), another Toeplitz matrix, generated by the impulse response \( h' \), whose frequency response is \( H' (\omega) \). It should be pointed out, however, that the analysis in this paper can be easily carried out also for the case of mismatch in the input distribution, or mismatch in the noise distribution, which has been already considered in [1]. In the matched case, \( P = P' \), we use the notation \( \text{mse} (X | Y) = \text{mse} (X | Y) \).

A very important function, which is pivotal to the derivation of both the estimator and the MSE is the partition function, defined as follows.

**Definition 1 (Partition Function)** Let \( \lambda = (\lambda_1, \ldots, \lambda_n) \) be a column vector of \( n \) real-valued parameters. The partition function w.r.t. the joint distribution \( P (x, y) \), denoted by \( Z (y, \lambda) \), is defined as

\[
Z (y, \lambda) = \sum_{x \in C} \exp \{ \lambda^T x \} P (x, y).
\]

Accordingly, under the above described model, we have that

\[
P' (y | x) = \frac{1}{(2\pi/\beta)^{n/2}} \exp \left[ -\beta \| y - A' x \|^2 / 2 \right].
\]

and so, the mismatched partition function is given by

\[
Z' (y, \lambda) = \sum_{x \in C} \exp \{ \lambda^T x \} P' (x, y)
\]

\[
= (2\pi/\beta)^{-n/2} \sum_{x \in C} e^{-nR} \exp \left[ -\beta \| y - A' x \|^2 / 2 + \lambda^T x \right].
\]

The role of \( \lambda \) in the above partition function is in computing the conditional mean estimator and the MSE. Indeed, it is easy to see that the gradient of \( \ln Z^* (y, \lambda) \) w.r.t. \( \lambda \), computed at \( \lambda = 0 \), simply gives the conditional mean estimator, i.e.,

\[
E' (X | Y = y) = \nabla_{\lambda} \ln Z^* (y, \lambda) |_{\lambda=0}
\]

where \( \nabla_{\lambda} \) denotes the gradient operator w.r.t. \( \lambda \). Also, in the matched case, it can be verified that the expectation of the Hessian of \( \ln Z (y, \lambda) \) w.r.t. \( \lambda \), computed at \( \lambda = 0 \), gives the MMSE, i.e.,

\[
\text{mmse} (X | Y) = \text{tr} \left\{ \nabla^2_{\lambda} \ln Z (y, \lambda) |_{\lambda=0} \right\}
\]

where \( \nabla^2_{\lambda} \) denotes the Hessian operator w.r.t. \( \lambda \). Using (7), the mismatched MSE can be calculated as

\[
\text{mse} (X | Y) = \sum_{i=1}^n \left\{ (X_i - E' (X_i | Y))^2 \right\}
\]

\[
= n \left\{ E (X_i^2) + E \left\{ \left[ E' (X_i | Y) \right] \right\}^2 \right\}
\]

All the above relations (and further) can be found in [1].

### III. Main Result and Discussion

In this section, our main results are presented and discussed. Due to space limitation, the proofs of all the following results are omitted and can be found in [5]. The asymptotic MMSE is given in the following theorem.

**Theorem 1 (Asymptotic MMSE)** Consider the model defined in Section II, and assume that the sequence \( \{ h_k \} \) is square summable. Then, the asymptotic MMSE is given by

\[
\lim_{n \to \infty} \text{mmse} (X | Y) = \left\{ \frac{1}{n} \int_0^n P_{\omega} P_{\omega} d\omega \right\}, \quad R > R_c
\]

\[
\leq R_c
\]

where \( R_c = \frac{1}{\pi} \int_0^n \frac{P_{\omega} P_{\omega} d\omega}{1 + |H (\omega)|^2} \)

\[
\text{lim}_{n \to \infty} \text{mmse} (X | Y) = \end{align}

From the above result, it can be seen that for \( R < R_c \), the MMSE essentially vanishes since the correct codeword can be reliably decoded, whereas for \( R > R_c \), the MMSE is simply the estimation error which results by the Wiener filter that would have been applied had the input been a zero-mean, i.i.d. Gaussian process, with variance \( 1/\beta \). Accordingly, it can be shown that (as a byproduct of the analysis) the MMSE estimator is exactly the Wiener filter. In the jargon of statistical mechanics of spin arrays (see for example [4, Ch. 6]), the range of rates \( R \leq R_c \), correspond to the ordered
phase (or ferromagnetic phase) in which the partition function is dominated by the correct codeword (and hence so is the posterior), while the range of rates \( R > R_c \) corresponds to the paramagnetic phase, in which the partition function is dominated by an exponential number of wrong codewords.

In contrast to the MMSE, unfortunately, the mismatched MSE does not lend itself to a simple closed-form expression. This complexity stems from the complicated dependence of the partition function on \( \lambda \). Nevertheless, despite of the non-trivial expressions, it should be emphasized that the obtained MSE expression has a single letter formula, and thus, practically, it can be easily calculated at least numerically. Due to the complicated expressions obtained for the MSE, in the following, we only present the general structure/behavior (in the sense of phase transitions) of the MSE without presenting the absolute error itself. It is shown in [5] that the MSE takes the following form: For the absolute error itself. It is shown in [5] that the MSE takes the following form: For 

\[
\lim_{n \to \infty} \text{mse}(X | Y) = \begin{cases} 
0, & R \leq R_e \\
E_p, & R > R_e 
\end{cases} 
\]  

(11)

and for \( R < 0 \) it is given by

\[
\lim_{n \to \infty} \text{mse}(X | Y) = \begin{cases} 
0, & R \leq R_f \\
E_g, & R_f < R \leq R_e \\
E_p, & R > R_e 
\end{cases} 
\]  

(12)

where the various parameters \( (R_d, R_e, \text{etc.}) \) in the above expressions are not presented here due to space limitations, but can be found in [5]. Thus, it can be seen that in the mismatched case, there is additional intermediate range (when \( R_d < 0 \)), which in statistical mechanics jargon is analogous to the glassy phase (or “frozen” phase), in which the partition function is dominated by a sub-exponential number of wrong codewords. Intuitively, in this range, we may have the illusion that there is relatively little uncertainty about the transmitted codeword, but this is wrong due to the mismatch (as the main support of the mismatched posterior belongs to incorrect codewords). In Section IV, we will relate each one of the two cases \( R_d \geq 0 \) and \( R_d < 0 \), to “pessimistic” and “optimistic” behaviors of the receiver, which were already mentioned in the Introduction.

In the following, we state a few general qualitative properties of the various quantities appearing in the obtained results. Similarly to [1], it turns out that the absolute error \( E_p \) is independent on \( R \), while \( E_g \) depends on \( R \) non-trivially. Accordingly, unlike the matched and pessimistic mismatched cases, the MSE is not piecewise constant in the whole range of rates when the estimator is optimistic. Also, as the SNR increases, the absolute errors \( E_g \) and \( E_p \) decrease, while the critical ferromagnetic rate \( R_e \) (if \( R_d \geq 0 \) and \( R_f \) otherwise) increases, as should be expected. Finally, while in the matched case the MMSE is independent of the filter/channel phase (readily seen from Theorem 1), in the mismatched case, this conclusion is not true anymore. This fact is demonstrated in Section IV.

Finally, note that it is tempting to think that there should not be a range of rates for which the MSE (MMSE) vanishes, as we deal with an estimation problem rather than a decoding problem. Nonetheless, since codewords are being estimated, and there are a finite number of them, for low enough rates (up to some critical rate) the posterior is dominated by the correct codeword, and thus asymptotically, the estimation can be regarded as a maximum a posteriori probability (MAP) estimation, and so the error vanishes. In the same breath, note that this is not the case if mismatch in the input distribution is considered. For example, if the receiver’s assumption on the transmitted energy is wrong, then no matter how low the rate is, there will always be an inherent error which stems from the fallacious averaging over a hypersphere with wrong radius (wrong codebook). Precisely, in this case, the estimated codeword will differ from the real one by an inevitable scaling of \( \sqrt{P_p}/P_s \), where \( P_p \) is the mismatched power.

**Remark 1** Although we have assumed that the transmitted codeword has a flat spectrum, the analysis can readily be extended to any input spectral density \( S_x(\omega) \).

### IV. Examples

In this section, we provide two examples in order to illustrate the theoretical results presented in the previous section. In particular, we present and explore the phase diagrams and the MSE’s as functions of the rate and some parameters of the mismatched channel. The main goal in these examples is further understanding of the role of the true and the mismatched probability measures in creating phase transitions. Further examples can be found in [5].

**Example 1** Let \( H(\omega) \) be a multiband filter given by

\[
H(\omega) = \begin{cases} 
1, & |\omega| \leq \frac{\omega_R}{8} \text{ or } |\omega| \leq \frac{\omega_L}{8}, \\
0, & \text{else}
\end{cases} 
\]  

(13)

and let the mismatched filter be given by a band-pass filter

\[
H'(\omega) = \begin{cases} 
1, & \omega_L \leq |\omega| \leq \omega_R, \\
0, & \text{else}
\end{cases} 
\]  

(14)

with constant bandwidth, \( \omega_R - \omega_L = \pi/8 \), i.e., smaller than the real one. In the numerical calculations, we again chose \( \beta = P_p = 1 \). Figures 1 and 2 show, respectively, the phase diagram and the MSE as functions of \( R \) and \( \omega_L \). First, observe that for \( \omega_R < \pi/4 \), which means that \( H'(\omega) \) and \( H(\omega) \) are equal to one over non intersecting frequency ranges, there is no ferromagnetic phase, as expected. Accordingly, for \( \omega_R > \pi/4 \), the ferromagnetic phase begins to play a role, and it can be seen that for \( \pi/4 < \omega_R < \pi/2 \), which means maximal intersection between the two filters, the range of rates for which the ferromagnetic phase dominates the partition function is maximal. Since the matched filter has two bands, obviously, the same behavior appears also in the second band. Thus, in this example, we actually obtain two disjoint glassy (and ferromagnetic) regions, which correspond to the two bands of the matched filter. Also, as shown in Fig. 2, in the ranges where no ferromagnetic phase exists, the MSE within the paramagnetic phase is larger than the MSE within the regions where ferromagnetic phase does exists, as one would expect.

**Remark 2** Example 1 essentially demonstrates that there can be arbitrarily many phase transitions. Generally speaking, for a matched multiband filter with \( N \) disjoint bands, and a
mismatched bandpass filter (with small enough bandwidth), there are $N$ disjoint glassy and ferromagnetic phases.

**Example 2** Let $H(z)$ be given by

$$H(z) = z - 2 \cos(0.8 \pi z^{-1}) - 1 - e^{-0.8 \pi z^{-1}}$$

and let the mismatched filter be given as

$$H'(z) = H(z) \cdot z^{-d}$$

where $d \in \mathbb{Z}$ is a mismatched delay. As before, in the numerical calculations, we chose $\beta = P_e = 1$. Figures 3 and 4 show, respectively, the phase diagram and the MSE as functions of $R$ and $d$. First, we see that $R_{gw}$ is constant, which makes sense since it can be shown that $R_e$ is independent of the delay [5]. Also, for all $d \neq 0$ there is a glassy phase, which means that for all $d \neq 0$, $R_d \leq 0$. More importantly, it can be observed that the MSE vanishes (or equivalently, the ferromagnetic phase dominates the partition function) only in the case that $d = 0$, namely, zero delay. This is a reasonable result, as a delay of one sample (linear phase) is enough to cause a serious degradation in the MSE. Actually, for any fixed rate the error is constant, independently of the delay, due to the fact that the MSE takes into account all the possible codewords in the codebook. Finally, note that the MSE is larger in the glassy region than in the paramagnetic region\(^1\). This is also a reasonable result: As the rate increases, and hence more codewords are possible, since the MSE estimator is actually a weighted average (w.r.t. the posterior) over the codewords, the MSE can only decrease (each codeword in the codebook contributes approximately the same estimation error). Accordingly, for small codebooks (low rates) the MSE is larger, since the averaging is performed over “fewer” codewords.

**ACKNOWLEDGMENT**

This research was partially supported by The Israeli Science Foundation (ISF), grant no. 412/12.

**REFERENCES**


\(^1\) Note that the MSE, in contrast to the MMSE, must not be monotonically increasing as a function of the rate.