Weak–Noise Modulation–Estimation of Vector Parameters*

Neri Merhav

The Andrew & Erna Viterbi Faculty of Electrical Engineering
Technion - Israel Institute of Technology
Technion City, Haifa 32000, ISRAEL
E–mail: merhav@ee.technion.ac.il

Abstract

We address the problem of modulating a parameter onto a power–limited signal, transmitted over a discrete–time Gaussian channel and estimating this parameter at the receiver. Continuing an earlier work, where the optimal trade–off between the weak–noise estimation performance and the outage probability (threshold–effect breakdown) was studied for a single (scalar) parameter, here we extend the derivation of the weak–noise estimation performance to the case of a multi–dimensional vector parameter. This turns out to be a non–trivial extension, that provides a few insights and it has some interesting implications, which are discussed in depth. Several modifications and extensions of the basic setup are also studied and discussed.

Index Terms: parameter estimation, threshold effect, modulation, waveform communication, channel capacity, quantization.

*This research was supported by the Israel Science Foundation (ISF), grant no. 137/18.

1 Introduction

We consider the scenario of communicating a real–valued parameter u by using an additive white Gaussian noise (AWGN) channel n times, that is,

$$y_t = x_t + z_t, \qquad t = 1, 2, \ldots, n, \tag{1}$$

where x_t is the t–th coordinate of a channel input vector, $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) = f_n(u)$, which depends on u, and which is limited by a power constraint, $\|\boldsymbol{x}\|^2 \le nP$, $\{z_t\}$ are independent, zero–mean, Gaussian random variables with variance σ^2, and y_t is the t–th component of the channel output vector, $\boldsymbol{y} = (y_1, y_2, \ldots, y_n)$. In the simplest case, the parameter is assumed to take on values within a finite interval, say, $[0, 1]$. Generally speaking, our main interest is in the following question: what is the best one can do in estimating u from $\boldsymbol{y}$, given that there is freedom to design, not only the estimator at the receiver, but also the modulation function, i.e., the choice of $f_n(\cdot)$? How rapidly can the estimation error decay as n grows without bound, given that the optimal modulator and estimator are used?

This problem is actually the discrete–time counterpart of the classical "waveform communication" problem (in the terminology of [15, Chap. 8]), and it can be approached either from the information–theoretic perspective or the estimation–theoretic perspective. Viewed from the information–theoretic viewpoint, this is a joint source–channel coding problem (see, e.g., [8] and references therein), where one source symbol u is conveyed by n channel uses, and it also belongs to the realm of Shannon–Kotel'nikov mappings (see, e.g., [5], [6], [7], [9] and references therein). From the estimation–theoretic perspective, every given modulator $f_n(\cdot)$ dictates a certain parametric family of conditional densities of $\boldsymbol{y}$ given u, for which estimation theory provides an arsenal of lower bounds on the estimation performance (most notably, in terms of the mean square error), along with good estimators, e.g., the maximum likelihood (ML) estimator in the non–Bayesian case, the maximum a–posteriori (MAP) estimator and the conditional mean, in the Bayesian case, and many others.

Conceptually, the simplest form of modulation is linear, that is, the case where $f_n(u) = u \cdot \boldsymbol{s}$ for some fixed vector $\boldsymbol{s}$ that is independent of u. Here, the ML estimator achieves the non–Bayesian Cramér–Rao lower bound (CRLB) for unbiased estimators and the conditional mean estimator is, of course, optimal in the Bayesian case. However, the class of linear modulators is very limited and much better performance can be obtained by non–linear modulation, at least in the high signal–to–noise (SNR) limit. On the other hand, the inherent caveat of non–linear modulation is the well–known *threshold effect* (see, e.g., [15, Chap. 8]). The threshold effect is a phenomenon of an abrupt passage between two very different types of behavior when the SNR crosses a certain critical level.

For high SNR, the mean square error (MSE) behaves essentially as in linear modulation, where it roughly achieves the CRLB. Below a certain SNR level, however, the performance breaks down and the MSE increases very abruptly. As presented in [15, Chap. 8], for a given non–linear modulator, we are able to identify a certain anomalous error event, or *outage event*, whose probability becomes considerably large when the threshold is crossed.

Existing results on non–linear modulation and estimation include many performance bounds (see, e.g., [1], [2], [3] as well as references therein), with no distinction between errors associated with relatively weak noise and anomalous errors. In other words, both kinds of errors are weighted in the evaluation of the total MSE under equal footing. One might argue, however, that it would be very reasonable to make a distinction between these two types of errors with the rationale that estimation in the event of outage is actually meaningless. In fact, some analysis in the spirit of such a separation appears already in [15, pp. 661–674], but not quite in a very formal framework.

A more systematic study along the line of separating weak–noise errors from anomalous errors, appears in [8, Section IV.A], where the communication system design problem was formulated as a constrained optimization problem of a transmitter and receiver, so as to minimize the conditional MSE given that no outage event has occurred, under the constraint that the outage probability would be kept below a prescribed (small) constant. In this problem, the definition of the outage event was left as a degree of freedom to be optimized, in addition to the optimization of the modulator and the estimator. It was argued in [8], that the data–processing lower bound is asymptotically achieved by a simple system that is based on quantization of the parameter, followed by a good channel code, where on the receiver side, the digital message is decoded and then mapped back to the quantized parameter value. The outage event is then the error event of the channel coding part and the weak–noise MSE is just the quantization error. There seems to be a gap, however, between the achievability and the converse bound of [8], because the data–processing converse bound corresponds to a situation where there is no freedom to allow an outage event, whereas the weak–noise upper bound therein suppresses the contribution of outage event (by definition). On the other hand, conditioning on the non–outage event might create conditional dependence between the source and the channel, in which case the data processing theorem would no longer apply. Therefore, it is not clear that the data processing inequality is the best tool for the derivation of converse bounds for this problem.

More recently, in [13], the approach of [8] was sharpened in two ways: first, in both the converse bound and the achievability, an outage event was formally allowed, and its choice was subjected to optimization. In other words, the lower bound and the upper bound in [13] are consistent with each other as they are associated with the same setup. Secondly, instead of constraining the outage probability by a small constant, the outage probability constraint imposed in [13] referred to a given exponentially decaying function of n, i.e., $e^{-\lambda n}$ for some given constant $\lambda > 0$. Under this constraint, the fastest possible decay rate of the MSE, or more generally, the most rapid decrease of the expectation of any symmetric, convex function of the estimation error, was sought. Specifically, in [13], upper and lower bounds were derived, which coincide in the high SNR limit within a certain interval of small λ. The achievability scheme proposed therein was based on quantization and channel coding using lattice codes with Voronoi cells whose shape is very close to n–dimensional spheres [16, Chap. 7], for large n. This scheme is asymptotically optimal universally for a wide family of error cost functions.

This paper is a further development of [13]. Referring to the regime where the outage probability is merely required to tend to zero, not necessarily exponentially in n (i.e., $\lambda > 0$ is arbitrarily small), it extends the results of [13] from the single (scalar) parameter to a multi–dimensional vector parameter, modulating a power–limited signal. The reason for focusing on the case where λ is arbitrarily small ($\lambda \to 0$) will be explained in the sequel.

This extension turns out to be rather non–trivial. To understand the reason, the following background is important: the weak–noise lower bound of the one–dimensional parameter [13] depends on the modulator only via the length of the signal locus curve,

$$L(f_n) = \int_0^1 \left\| \frac{\partial f_n(u)}{\partial u} \right\| \mathrm{d}u, \tag{2}$$

and then, by deriving a universal upper bound to $L(f_n)$, the converse bound is obtained. It is then natural to expect, at least at first thought, that when it comes to higher dimensions, the role of the one–dimensional signal locus length in the lower bound, would be replaced by the hyper–area/volume of the signal manifold in the higher dimension. For example, if the transmitted vector, $f_n(u,v)$, depends on a two–dimensional vector parameter, $(u,v) \in [0,1]^2$, it is appealing to expect, in view of [13], that the lower bound would now depend on the surface area of the two–dimensional signal manifold, which is given by[1]

$$S(f_n) = \int_0^1 \int_0^1 \sqrt{\left\| \frac{\partial f_n(u,v)}{\partial u} \right\|^2 \cdot \left\| \frac{\partial f_n(u,v)}{\partial v} \right\|^2 - \left\langle \frac{\partial f_n(u,v)}{\partial u}, \frac{\partial f_n(u,v)}{\partial v} \right\rangle^2} \mathrm{d}u \mathrm{d}v, \tag{3}$$

and then the converse would be obtained from an upper bound on $S(f_n)$ (see, e.g., [5, Section 2.5]).

Unfortunately, we have not been able to obtain lower bounds that depend the modulator solely via $S(f_n)$, even in the two–dimensional case, let alone the parallel quantities in higher dimensions.

[1]The integrand of the expression of $S(f_n)$ can be thought of as the product $\|\partial f_n(u,v)/\partial u\| \cdot \|\partial f_n(u,v)/\partial v\| \cdot |\sin\theta(u,v)|$, where $|\sin\theta(u,v)| = \sqrt{1-\cos^2\theta(u,v)}$, $\theta(u,v)$ being the angle between the vectors $\partial f_n(u,v)/\partial u$ and $\partial f_n(u,v)/\partial v$. This product is equal to the area of the infinitesimal parallelogram defined by the vectors $\mathrm{d}u \cdot \partial f_n(u,v)/\partial u$ and $\mathrm{d}v \cdot \partial f_n(u,v)/\partial v$ with the angle of $\theta(u,v)$ in between.

Fortunately enough, however, it turns out that it is possible to bypass the need to handle such higher dimensional objects. The idea is to use one–dimensional paths that "scan" (high–resolution grids of) the higher dimensional signal manifold with a carefully chosen density, and then one can basically use the one–dimensional derivations of [13]. But different types of scans may yield different lower bounds, and then the question is what is the best scanning strategy that would yield a tight (achievable) lower bound. It turns out that it is best to exhaust the parameter space along "diagonal" straight lines. For example, in the two dimensional case of (u, v), the idea is to "scan" along straight lines with a slope of 45 degrees, that is, the family of lines $u - v = c$ in the (u, v)–plane, for all possible values of c with the right resolution. The matching achievability result will be based on simple uniform quantization and channel coding, as in [13] and some earlier works. Here too, this scheme is asymptotically optimal universally for a wide class of error cost functions.

More details on both the converse and the achievability will be provided, of course, in the sequel, but one of the interesting conclusions from our results is that in optimal parametric modulation, the various components of the parameter vector share the same resources of channel capacity, and therefore there is an inherent trade–off among the achievable accuracies of their estimation: if more estimation accuracy is given to one component of the parameter vector, then it comes at the expense of the accuracy that can be obtained in the other components. A similar behavior was observed also in earlier studies under different regimes and different performance metrics, such as the large deviations performance [11], general moments of the estimation error [12], and the mean square error in a multiple access modulation–estimation scenario [14]. This fact is not quite trivial if one takes into account that in many parameter estimation problems, there is no conflict and no interaction between the estimation errors achieved for the various components of the vector parameter. For instance, consider the case where the ML estimator is Fisher–efficient in a parametric model whose CRLB matrix is diagonal, like the example where the desired signal is a linear combination of several orthogonal basis functions, corrupted by Gaussian white noise, and the parameters are the coefficients of this linear combination. Here, the mean square error of each coefficient is not affected by the lack of knowledge of the other coefficients. The above described conclusion of this work tells us then that, nevertheless, whenever the modulator is also subjected to optimization, the resulting parametric family would exhibit interactions and trade–offs among the various parameters. On a related note, one of the conclusions from our main result is that the best trade-off is obtained when all components of the vector parameter contribute essentially evenly to the total error cost.

We also discuss a few extended versions of our basic communication model. One of them is modulation–estimation over the dirty–paper channel, where estimation performance is essentially unaffected by the lack of channel state information at the receiver. In another model of unknown channel interference, we discuss how universal decoding techniques can be harnessed for the purpose of universal estimation. Finally, we demonstrate how constraints on the modulated signal structure (which limit the freedom in the signal design), may fundamentally affect performance, by showing that larger (tighter) lower bounds may apply. This fact has implications, for example, in multiple access communication.

The outline of the remaining part of the paper is as follows. In Section 2, we establish notation conventions, define the problem more formally, and spell out our assumptions. In Section 3, we present the main result and discuss its insights, implications and a few variations. Section 4 is devoted to the proof of the main theorem. Finally, in Section 5, we summarize our main conclusions and mention a few possible further questions for future research.

2 Notation, Problem Formulation and Assumptions

2.1 Notation

Throughout the paper, random variables will be denoted by capital letters and specific values they may take will be denoted by the corresponding lower case letters. Random vectors and their realizations will be denoted, respectively, by capital letters and the corresponding lower case letters, both in the bold face font. For example, the random vector $\boldsymbol{X} = (X_1, \ldots, X_n)$, ($n$ – positive integer) may take a specific vector value $\boldsymbol{x} = (x_1, \ldots, x_n)$. The probability of an event $\mathcal{E}$ will be denoted by $\Pr\{\mathcal{E}\}$, and the expectation operator will be denoted by $\boldsymbol{E}\{\cdot\}$. For two positive sequences a_n and b_n, the notation $a_n \doteq b_n$ will stand for equality in the exponential scale, that is, $\lim_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} = 0$. Similarly, $a_n \stackrel{\cdot}{\le} b_n$ means that $\limsup_{n\to\infty} \frac{1}{n} \log \frac{a_n}{b_n} \le 0$, and so on. The indicator function of an event $\mathcal{E}$ will be denoted by $\mathcal{I}\{E\}$. Finally, the notation $[x]_+$ will stand for $\max\{0, x\}$.

2.2 Problem Formulation and Assumptions

The problem formulation is similar to that of [13], but with a few modifications that accommodate the extensions being addressed. We consider the following communication system model. The transmitter needs to communicate to the receiver, a given value of a vector parameter $\boldsymbol{u} = (u_1, \ldots, u_d) \in [0,1]^d$, and to this end, it uses the channel n times, subject to a given power constraint. The receiver has to estimate $\boldsymbol{u}$ from the n noisy channel outputs. More precisely, given $\boldsymbol{u} \in [0,1]^d$, the transmitter outputs a vector $\boldsymbol{x} = f_n(\boldsymbol{u}) \in \mathbb{R}^n$, subjected to the power constraint,

$\|\boldsymbol{x}\|^2 \leq nP$, P being the allowed power. The received vector is

$$\boldsymbol{Y} = f_n(\boldsymbol{u}) + \boldsymbol{Z}, \tag{4}$$

where $\boldsymbol{Z} \in \mathbb{R}^n$ is a zero–mean Gaussian noise vector with covariance matrix $\sigma^2 \cdot I$, I being the $n \times n$ identity matrix. The receiver employs an estimator $\hat{\boldsymbol{u}} = g_n[\boldsymbol{y}]$ (with $\boldsymbol{y}$ denoting a realization of the random vector $\boldsymbol{Y}$) of the vector parameter $\boldsymbol{u}$, where $g_n : \mathbb{R}^n \to [0,1]^d$. For every $u \in [0,1]^d$, let $\mathcal{O}_n(\boldsymbol{u}) \subset \mathbb{R}^n$ denote an event defined in the space of noise vectors, $\{\boldsymbol{z}\}$, henceforth called the *outage event* (or the *anomalous error event*) given $\boldsymbol{u}$. In the sequel, we will also use the notation

$$\mathcal{Y}_n(\boldsymbol{u}) \stackrel{\triangle}{=} \mathcal{O}_n(\boldsymbol{u}) + f_n(\boldsymbol{u}) \equiv \{\boldsymbol{y} = f_n(\boldsymbol{u}) + \boldsymbol{z} : \ \boldsymbol{z} \in \mathcal{O}_n(\boldsymbol{u})\}, \tag{5}$$

and similarly,

$$\mathcal{Y}_n^{\mathrm{c}}(\boldsymbol{u}) \stackrel{\triangle}{=} \mathcal{O}_n^{\mathrm{c}}(\boldsymbol{u}) + f_n(\boldsymbol{u}). \tag{6}$$

The goal is to propose a communication system, defined by a modulator, $f_n(\cdot)$, that complies with the power constraint, and a receiver, $g_n[\cdot]$, along with a family of outage events, $\mathcal{O}_n = \{\mathcal{O}_n(\boldsymbol{u}), \ \boldsymbol{u} \in [0,1]^d\}$, in order to minimize

$$\sup_{\boldsymbol{u} \in [0,1]^d} \boldsymbol{E}\left\{\rho(g_n[f_n(\boldsymbol{u}) + \boldsymbol{Z}] - \boldsymbol{u}) \middle| \boldsymbol{Z} \in \mathcal{O}_n^{\mathrm{c}}(\boldsymbol{u})\right\} \tag{7}$$

under the constraint

$$\sup_{\boldsymbol{u} \in [0,1]^d} \Pr\{\mathcal{O}_n(\boldsymbol{u})\} \leq \delta_n \to 0, \tag{8}$$

where the expectation $\boldsymbol{E}\{\cdot\}$ in (7) and the probability $\Pr\{\cdot\}$ in (8) are with respect to (w.r.t.) the randomness of $\boldsymbol{Z}$. Here, δ_n is a positive sequence that tends to zero (not necessarily exponentially fast as in [13]), $\mathcal{O}_n^{\mathrm{c}}(\boldsymbol{u})$ is the complement of $\mathcal{O}_n(\boldsymbol{u})$, and $\rho : [-1,1]^d \to \mathbb{R}^+$ is referred to as the *error cost function* (ECF), which is assumed to be a weighted L_q distance function, i.e.,

$$\rho(\boldsymbol{\epsilon}) = \rho(\epsilon_1, \ldots, \epsilon_d) = \sum_{i=1}^{d} W_i \cdot |\epsilon_i|^q, \tag{9}$$

where power $q \geq 1$ and the weights, $W_i \geq 0$, $i = 1, \ldots, d$, are given constants. The weights reflect the relative importance of good estimation of the various components of $\boldsymbol{u}$. They are constants in the sense that they are independent of $\boldsymbol{\epsilon}$, but they are allowed to depend on n. Since we focus on the asymptotic performance in the exponential scale, it is reasonable to choose $\{W_i\}$ to be exponential functions of n, as otherwise, they their particular choice will have no effect on the exponential order of the results. In particular, we set $W_i = e^{-na_i}$, $i = 1, 2, \ldots, d$, for some real–valued constants, $a_1, \ldots, a_d$, that are independent of n. Thus, eq. (9) now reads

$$\rho(\boldsymbol{\epsilon}) = \rho(\epsilon_1, \ldots, \epsilon_d) = \sum_{i=1}^{d} e^{-na_i} \cdot |\epsilon_i|^q. \tag{10}$$

Let $\mathcal{C}_n$ denote the class of all families of outage events, $\{\mathcal{O}_n\}$, that satisfy (8).

As in [13], we will be interested in characterizing the fastest possible exponential decay rate of (7) subject to (8), where δ_n is allowed to decay arbitrarily slowly (or even remain fixed, independently of n). As mentioned in the Introduction, this is different from [13], where δ_n was taken to be an exponential function of n, i.e., $e^{-\lambda n}$ for a given $\lambda > 0$. Our present setup then corresponds to the limit $\lambda \downarrow 0$ of [13].[2]

More formally, given sequences of modulators $F = \{f_n(\cdot)\}_{n\geq 1}$ (all satisfying the power constraint), estimators $G = \{g_n[\cdot]\}_{n\geq 1}$, and families of outage events, $\mathcal{O} = \{\mathcal{O}_n\}_{n\geq 1}$ (with $\mathcal{O}_n \in \mathcal{C}_n$), let

$$\mathcal{E}(F,G,\mathcal{O}) = \liminf_{n\to\infty}\left[-\frac{1}{n}\ln\left(\sup_{\boldsymbol{u}\in[0,1]^d}\boldsymbol{E}\left\{\rho(g_n[f_n(\boldsymbol{u})+\boldsymbol{Z}]-\boldsymbol{u})\middle|\boldsymbol{Z}\in\mathcal{O}_n^{\mathrm{c}}(\boldsymbol{u})\right\}\right)\right]. \tag{11}$$

Definition 1 *We say that E is an achievable weak–noise error cost exponent if there exists a sequence of communication systems $(F,G,\mathcal{O})$, such that the following two conditions hold at the same time:*

(a) $\mathcal{E}(F,G,\mathcal{O}) \geq E$, and

(b) For any $\boldsymbol{u}_1, \boldsymbol{u}_2 \in [0,1]^d$ and any $E' < E$, if $\rho(\boldsymbol{u}_1 - \boldsymbol{u}_2) \stackrel{\cdot}{\geq} e^{-nE'}$ then $\mathcal{Y}_n^{\mathrm{c}}(\boldsymbol{u}_1)\cap\mathcal{Y}_n^{\mathrm{c}}(\boldsymbol{u}_2) = \emptyset$ for all sufficiently large n, uniformly in $\boldsymbol{u}_1, \boldsymbol{u}_2 \in [0,1]^d$.

The rationale behind these two conditions is that they formally describe the distinction between the weak–noise mode (part (a)) and the outage mode (part (b)). If the weak–noise error cost is essentially no larger than e^{-nE} (in terms of the exponential order), then two parameter values whose distance (in the sense of the distance function ρ) is exponentially strictly larger than e^{-nE}, should be far apart also in the signal domain, so that they would not be confused as long as the noise is not exceptionally strong.

Our purpose is to derive upper and lower bounds to the largest achievable weak–noise error cost exponent as a function of the SNR, $\gamma = P/\sigma^2$, which will be denoted by $E(\gamma)$. We also denote by $C(\gamma)$ the capacity of the AWGN with SNR γ, that is,

$$C(\gamma) = \frac{1}{2}\ln(1+\gamma). \tag{12}$$

[2]There are two motivations for focusing on the case $\lambda \downarrow 0$. The first is that it is coherent with the regimes described in [15, Chap. 8] and in [8], and the second is that it allows full compatibility between the achievable performance and the converse bound, unlike the case of a general value of λ, where the gap was closed only in the high SNR limit and only for a certain range of values of λ. Also, for a general positive λ, in the achievability scheme of [13] only a good, capacity–achieving lattice code must have been used, whereas here, any capacity–achieving channel code can be used.

Finally, the following assumptions will be made.

A.1 The parameters of the ECF satisfy[3]

$$d \cdot \max_{1\le i\le d} a_i - \sum_{j=1}^{d} a_j \le qC(\gamma). \tag{13}$$

A.2 Consider any curve in the parameter space, i.e., $\mathcal{L} = \{\boldsymbol{u}(t) = [u_1(t), \ldots, u_d(t)] :\ 0 \le t \le 1\}$, where $u_i(\cdot)$ are arbitrary continuous functions, and let $\{f_n[\boldsymbol{u}(t)],\ 0 \le t \le 1\}$ be the image of this curve in the signal space. For every positive integer m, let $0 = t_0^m < t_1^m < t_2^m < \ldots t_m^m < t_{m+1}^m = 1$ be a partition of the unit interval. It is assumed that if $\max_{1\le \ell \le m+1}[t_\ell^m - t_{\ell-1}^m] \to 0$ as $m \to \infty$, then $\sum_{\ell=1}^{m+1} \|f_n[\boldsymbol{u}(t_\ell)] - f_n[\boldsymbol{u}(t_{\ell-1})]\|$ tends to a limit[4] that is independent of the particular sequence of partitions.

A.3 Definition 1 applies to the communication system $(F, G, \mathcal{O})$ for some $E > 0$.

3 Main Result

Let us define

$$E(\gamma) \stackrel{\Delta}{=} \sup_{\{(R_1,\ldots,R_d):\ R_1+\ldots+R_d \le C(\gamma)\}} \min_{1\le i\le d} [a_i + qR_i]. \tag{14}$$

It is easy to show that under Assumption A.1, an equivalent expresion for $E(\gamma)$ is given by

$$E(\gamma) = \frac{qC(\gamma) + \sum_{i=1}^{d} a_i}{d}. \tag{15}$$

To see why this is true, observe that on the one hand, $E(\gamma)$ is upper bounded by

$$\begin{aligned} E(\gamma) &= \sup_{\{(R_1,\ldots,R_d):\ \sum_i R_i \le C(\gamma)\}} \min_{1\le i\le d} (a_i + qR_i) \\ &\le \sup_{\{(R_1,\ldots,R_d):\ \sum_i R_i \le C(\gamma)\}} \frac{1}{d}\sum_{i=1}^{d}(a_i + qR_i) \\ &= \frac{\sum_{i=1}^{d} a_i + qC(\gamma)}{d}, \end{aligned} \tag{16}$$

but on the other hand, equality is achieved by the assignment

$$R_i^* = \frac{C(\gamma)}{d} + \frac{1}{q}\left(\frac{1}{d}\sum_{j=1}^{d} a_j - a_i\right), \qquad i = 1, 2, \ldots, d. \tag{17}$$

[3] Since the l.h.s. of eq. (13) depends only on the parameters of the ECF and the r.h.s. depends only on the SNR, this can also be viewed as an assumption of a sufficiently high SNR. In fact, since only the ratios between the weights $\{W_i\}$ matter, this means that only the differences among $\{a_i\}$ are important, and then it could be convenient to shift them all such that $\sum_{i=1}^{d} a_i = 0$, in which case (13) would simplify to $d \cdot \max_{1\le i\le d} a_i \le qC(\gamma)$.

[4] This limit is the length of the signal locus curve, $\{f_n[\boldsymbol{u}(t)],\ 0 \le t \le 1\}$.

Assumption A.1 ensures that $R_i^* \geq 0$ for all $i = 1, \ldots, d$, a condition that will be required both in the direct part and the converse part of our main result, which is the following.

Theorem 1 *Consider the setup defined in Section 2 along with assumptions A.1 – A.3. Then, for every sequence of modulators (all satisfying the power constraint) receivers, and families of outage events $\mathcal{O}$ (with $\mathcal{O}_n \in \mathcal{C}_n$), the largest achievable weak–noise error cost exponent is $E(\gamma)$, given in (15).*

The proof of Theorem 1 appears in Section 4. The remaining part of this section is devoted to a discussion concerning the insights and the implications associated with this theorem.

The achievability. As is shown in the achievability part of the proof of Theorem 1, $E(\gamma)$ can be achieved using the very simple idea of source coding followed by capacity–achieving channel coding, that is, separate source– and channel coding (see also [13] and references therein, for the use of the same approach in somewhat different scenarios). Moreover, the source coding part does not even require any sophisticated vector quantization, but simply a uniform scalar quantization at rate R_i^* for each component u_i of $\boldsymbol{u}$ separately (see also [11] and [12]). Also, this approach is universally optimal no matter what the parameters of the error cost function, $\{a_i\}$ and q, may be, as long as the meet Assumption A.1. The only thing that depends on these parameters is the optimal assignment of the various quantization rates, $R_1^*, \ldots, R_d^*$. It is also interesting to note that the optimal rates are chosen such that all components of the vector parameter contribute essentially evenly to the total expected weak–noise error cost.

The converse. Since the achievability part of Theorem 1, namely, the inequality $\sup_{F,G,\mathcal{O}} \mathcal{E}(F, G, \mathcal{O}) \geq E(\gamma)$, is conceptually simple as described above (and with some more detail in the proof), the deeper and more interesting part of Theorem 1 is the converse part, which is the inequality $\sup_{F,G,\mathcal{O}} \mathcal{E}(F, G, \mathcal{O}) \leq E(\gamma)$. This is because it applies, of course, to any arbitrary modulation–estimation system, not only systems that are based on separate source– and channel coding. The converse part also establishes the fact the problem of modulation and estimation is fundamentally intimately related to channel coding and channel capacity.

Performance as a function of d. We observe from (15) that for a fixed value of $\sum_{i=1}^d a_i$ (for example, the case where $\sum_i a_i = 0$), the best achievable weak–noise error cost exponent is inversely proportional to the dimension d of the parameter vector. This behavior was observed also in [11] and [12], although different performance metrics were considered in those works. It is also anticipated even by the data processing bound pertaining to the Bayesian regime, where the components of the parameter vector are d i.i.d. random variables, and the resulting lower bound on the distortion depends (exponentially) on $C(\gamma)/d$. However, the data processing theorem does not yield a tight lower bound for this problem, because it lower bounds the total expected error cost, not just the weak–noise expected error cost.

Variations on the model. The observed intimate relationship between modulation–estimation and channel coding naturally triggers the consideration of other, more general communication scenarios, where results from the Shannon theory can easily be "imported".

1. The dirty–paper channel. One example is modulation–estimation across the dirty–paper coding model [4], whose capacity is well known to be the same as without interference. In particular, suppose that the channel model is now given by

$$y_t = x_t + s_t + z_t, \qquad t = 1, 2, \ldots, \tag{18}$$

where s_t is the t–th component of a random interference signal $\boldsymbol{s} = (s_1, \ldots, s_n)$ (with an arbitrary distribution), known non–causally to the transmitter only, z_t is Gaussian noise as before, and x_t is the t–th component of $\boldsymbol{x} = f_n(\boldsymbol{u}, \boldsymbol{s})$, which is subjected to the power constraint. The result of Theorem 1 remains intact for this model. To see why this is true, observe that in the direct part, one may apply the same scalar quantization as before, but replace the ordinary channel code by a capacity–achieving dirty–paper code [4]. In the converse part, assume a genie–aided receiver that has access to $\boldsymbol{s}$, and then apply the converse proof of Theorem 1 for every given (fixed) $\boldsymbol{s}$ (with the tube–packing argument therein applying to spheres centered at $\boldsymbol{s}$, instead of the origin).

2. Universal decoding at the service of universal estimation. As another example, consider again the channel (18), but with two differences relative to the dirty–paper setting discussed in the previous paragraph. The first is that the transmitter has no access to $\boldsymbol{s}$, namely, $\boldsymbol{x} = f_n(\boldsymbol{u})$ as before, and the second is that now $\boldsymbol{s}$ is no longer a random signal, but an unknown deterministic signal. Can we still achieve $E(\gamma)$ when neither the transmitter nor the receiver have access to $\boldsymbol{s}$? The answer turns out to be affirmative at least when $\boldsymbol{s}$ is known to have a certain structure. In particular, let us assume that

$$s_t = \sum_{i=1}^{\infty} \alpha_i \phi_{i,t}, \tag{19}$$

where $\{\phi_{i,t}\}$ is a given set of uniformly bounded basis functions (for example, sine and cosine functions), and the sequence of unknown coefficients, $\alpha_1, \alpha_2, \ldots$ is absolutely summable. In estimation–

theoretic terms, the parameters $\alpha_1, \alpha_2, \ldots$ are nuisance parameters, which cannot be controlled, as opposed to $\boldsymbol{u}$ which is the desired vector parameter, whose modulation is controlled by the signal design. Normally, the presence of unknown nuisance parameters causes degradation in the estimation performance of the desired parameters. Here, however, this is not the case, at least as far as weak–noise error cost exponents are concerned. In [10], a universal decoder for this channel was proposed and was shown to achieve the same random coding error exponent as the ML decoder that knows $\alpha_1, \alpha_2, \ldots$. A-fortiori, such a universal decoder achieves the capacity of this channel, $C(\gamma)$. Therefore, referring to the above–described achievability scheme for modulation and estimation, that is based on scalar quantization of each component of $\boldsymbol{u}$, followed by channel coding, if the receiver employs the universal decoder of [10], instead of the ML decoder, then the system still achieves $E(\gamma)$, in spite of the lack of knowledge of the (infinitely many) nuisance parameters. This means that universal decoding results can be harnessed for the purpose of universal estimation.

3. Tighter converse bounds for signals with a given structure. The purpose of our third and last example is to make the point that in the presence of structural constraints concerning the signal modulation, which means less freedom in the signal design, tighter converse bounds may sometimes be available. Consider, for example, the case of a two–dimensional vector parameter, $\boldsymbol{u} = (u_1, u_2)$, and a modulator that is enforced to have the form

$$f_n(u_1, u_2) = f_{n,1}(u_1) + f_{n,2}(u_2), \tag{20}$$

where $f_{n,1}(u_1)$ is limited to have power that does not exceed P_1 and $f_{n,2}(u_2)$ is limited to power P_2, and be orthogonal to $f_{n,1}(u_1)$. An immediate application of this model is the Gaussian multiple access channel (MAC), where two users wish to convey their parameters, u_1 and u_2 (see, e.g., [14]). Let $\gamma_1 = P_1/\sigma^2$ and $\gamma_2 = P_2/\sigma^2$ be the corresponding SNRs. For simplicity, let us consider the case $a_1 = a_2 = 0$. One converse bound for the weak–noise error cost exponent is obtained by a simple application of Theorem 1 for $d = 2$, regardless of the structure (20). This gives

$$E(\gamma_1 + \gamma_2) = \frac{qC(\gamma_1 + \gamma_2)}{2}. \tag{21}$$

Another lower bound on the weak–noise expectation of $|\epsilon_1|^q + |\epsilon_2|^q$ is obtained when each one of these two terms is treated individually, as if the corresponding parameter was the only unknown parameter (like in the case $d = 1$), whereas the other parameter was given (and then the signal associated with it could have been subtracted from $\boldsymbol{y}$). This would give $e^{-nqC(\gamma_1)} + e^{-nqC(\gamma_2)} \doteq e^{-nq\min\{C(\gamma_1), C(\gamma_2)\}}$. If one of the SNRs is relatively small, then this lower bound is tighter than the generic bound (21). On the other hand, if $\gamma_1 = \gamma_2$, the bound (21) is tighter.

4 Proof of Theorem 1

4.1 Achievability

The achievability of $E(\gamma)$ is easily established as follows. For a given, arbitrarily small $\epsilon > 0$, let $(R_1^*, \ldots, R_d^*)$ be as in (17), but with $C(\gamma)$ replaced by $C(\gamma) - \epsilon$. Now, uniformly quantize each component, u_i of $\boldsymbol{u}$ using $M_i = e^{nR_i^*}$ equally spaced quantization points with spacings of $1/M_i = e^{-nR_i^*}$ in between. Thus, the total set of quanitzed vectors forms a Cartesian grid of $\prod_{i=1}^d M_i = e^{n[C(\gamma)-\epsilon]}$ points. Each one of these quantized vectors is mapped into a codeword of a good channel code at rate $C(\gamma) - \epsilon$, which has a small maximum error probability, δ_n. If we define the decoding error event as the outage event $\mathcal{O}_n(\boldsymbol{u})$ of the communication system, then the weak–noise error cost is induced by the quantization error only. Since the absolute value of the quantization error, Δ_i, in u_i cannot exceed $e^{-nR_i^*}/2$, then

$$\rho(\Delta_1, \ldots, \Delta_d) \leq \sum_{i=1}^{d} e^{-na_i} \left| \frac{e^{-nR_i^*}}{2} \right|^q \doteq \exp\left\{-n \min_{1 \leq i \leq d} (a_i + qR_i^*)\right\} = e^{-n[E(\gamma) - q\epsilon/d]}. \tag{22}$$

Owing to the arbitrariness of ϵ, we can approach $E(\gamma)$ as closely as desired. Note that this achievability scheme essentially satisfies Definition 1, as it both achieves (arbitrarily closely) the weak–noise error cost of $E(\gamma)$ (which is part (a) of Definition 1), and it sends distant points in the parameter space to distant points in the signal space (part (b) of Definition 1): any two points, $\boldsymbol{u}_1$ and $\boldsymbol{u}_2$, whose distance is larger than the exponential order of $e^{nE'}$, with any $E' < E(\gamma) - q\epsilon/d$, must belong to different quantization cells in the parameter space, and hence be mapped onto different codewords, whose decoding regions, $\mathcal{Y}_n^c(\boldsymbol{u}_1)$ and $\mathcal{Y}_n^c(\boldsymbol{u}_2)$, in turn are disjoint.

4.2 Converse

For the sake of simplicity of the exposition, we confine attention to the case $d = 2$, with the understanding that the extension to a general dimension d will be straightforward. The proof is similar to that of the one–dimensional case of [13]. The difference is that instead of referring to a simple path of consecutive grid points along the interval $[0, 1]$, we define a two–dimensional grid and "scan" it along diagonal lines, as demonstrated in Fig. 1. To avoid cumbersome notation with many subscripts, let us denote the two components of the parameter vector by u and v, instead of u_1 and u_2. Similarly, a_1 and a_2 will be replaced by a_u and a_v, respectively. For two given positive integers, M_u and M_v (to be chosen later), consider the uniform grid of points $(u_i, v_j) \in [0, 1]^2$, where $u_i = i/M_u$, $i = 0, 1, \ldots, M_u - 1$, and $v_j = j/M_v$, $j = 0, 1, \ldots, M_v - 1$. Next, convert this two–dimensional grid into a one–dimensional sequence, $w_k = (u_i, v_j)$, $k = 0, 1, \ldots, M_u M_v - 1$, according to the following rule: for $\ell = (M_v - 1), (M_v - 2), \ldots, 1, 0, -1, \ldots, -(M_u - 2), -(M_u - 1)$, assign

$$w_{q(\ell)+i} = \left(\frac{i}{M_u}, \frac{i+\ell}{M_v}\right), \qquad i = [-\ell]_+, [-\ell]_+ + 1, \ldots, \min\{M_u, M_v - \ell\} - 1, \tag{23}$$

where $q(\ell)$ is the number of grid points strictly above the straight line $v = (u \cdot M_u + \ell)/M_v$ in the (u, v)–plane, and the operator $[\cdot]_+$ means positive truncation, i.e., $[t]_+ \stackrel{\Delta}{=} \max\{0, t\}$. This conversion into a one–dimensional sequence is demonstrated in Fig. 1 for $M_u = M_v = 4$, where u and v are designated by the horizontal and vertical axes, respectively. As can be seen, the scan begins at the top left corner, and then travels along all parallel diagonal lines (with a slope of 45 degrees), one–by–one, each one from the bottom–left to the top–right, until the entire grid is exhausted.

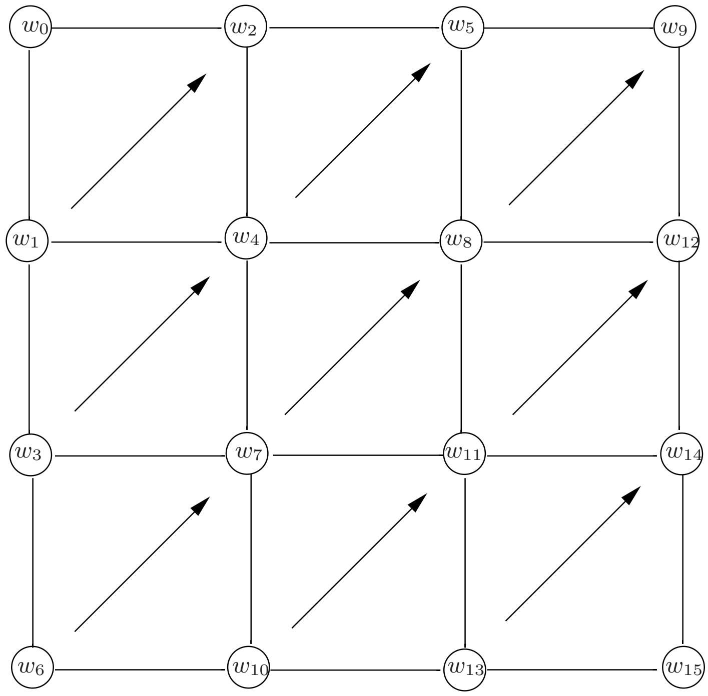

Figure 1: Scanning the two–dimensional grid along diagonal straight lines.

In the derivation below, the function $p(\cdot)$ designates the Gaussian density of the noise vector $\boldsymbol{z}$, that is,

$$p(\boldsymbol{z}) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{\|\boldsymbol{z}\|^2}{2\sigma^2}\right\}. \tag{24}$$

Consider now the following chain of inequalities, which is similar to the one in [13], but with a few twists.

$$
\begin{aligned}
& \sup_{(u,v)\in[0,1]^2} \boldsymbol{E}\left\{\rho(g_n[\boldsymbol{Y}]-(u,v))\middle|\boldsymbol{Z}\in\mathcal{O}_n^{\mathrm{c}}(u,v)\right\} \\
\geq\quad & \frac{1}{2M_uM_v}\sum_{k=0}^{M_uM_v-2}\Bigg[\boldsymbol{E}\Big(\rho(g_n[\boldsymbol{Y}]-w_k)\cdot\mathcal{I}\{\boldsymbol{Z}\in\mathcal{O}_n^{\mathrm{c}}(w_k)\}\Big)+ \\
& \boldsymbol{E}\left(\rho(g_n[\boldsymbol{Y}]-w_{k+1})\cdot\mathcal{I}\{\boldsymbol{Z}\in\mathcal{O}_n^{\mathrm{c}}(w_{k+1})\}\right)\Bigg] \\
\overset{(a)}{\geq}\quad & \frac{1}{2M_uM_v}\sum_{k\in\mathcal{S}}\int_{\mathcal{Y}_n^{\mathrm{c}}(w_k)\cap\mathcal{Y}_n^{\mathrm{c}}(w_{k+1})}[p(\boldsymbol{y}-f_n(w_k))\rho(g_n[\boldsymbol{y}]-w_k)+ \\
& p(\boldsymbol{y}-f_n(w_{k+1}))\rho(w_{k+1}-g_n[\boldsymbol{y}])]\,\mathrm{d}\boldsymbol{y} \\
\geq\quad & \frac{1}{M_uM_v}\sum_{k\in\mathcal{S}}\int_{\mathcal{Y}_n^{\mathrm{c}}(w_k)\cap\mathcal{Y}_n^{\mathrm{c}}(w_{k+1})}\min\{p(\boldsymbol{y}-f_n(w_k)),p(\boldsymbol{y}-f_n(w_{k+1}))\}\times \\
& \left[\frac{1}{2}\rho(g_n[\boldsymbol{y}]-w_k)+\frac{1}{2}\rho(w_{k+1}-g_n[\boldsymbol{y}])\right]\mathrm{d}\boldsymbol{y} \\
\overset{(b)}{\geq}\quad & \frac{1}{M_uM_v}\sum_{k\in\mathcal{S}}\int_{\mathcal{Y}_n^{\mathrm{c}}(w_k)\cap\mathcal{Y}_n^{\mathrm{c}}(w_{k+1})}\min\{p(\boldsymbol{y}-f_n(w_k)),p(\boldsymbol{y}-f_n(w_{k+1}))\}\times \\
& \rho\left(\frac{g_n[\boldsymbol{y}]-w_k}{2}+\frac{w_{k+1}-g_n[\boldsymbol{y}]}{2}\right)\mathrm{d}\boldsymbol{y} \\
=\quad & \frac{1}{M_uM_v}\sum_{k\in\mathcal{S}}\rho\left(\frac{w_{k+1}-w_k}{2}\right)\int_{\mathcal{Y}_n^{\mathrm{c}}(w_k)\cap\mathcal{Y}_n^{\mathrm{c}}(w_{k+1})}\min\{p(\boldsymbol{y}-f_n(w_k)),p(\boldsymbol{y}-f_n(w_{k+1}))\}\mathrm{d}\boldsymbol{y} \\
=\quad & \rho\left(\frac{1}{2M_u},\frac{1}{2M_v}\right)\cdot\frac{1}{M_uM_v}\sum_{k\in\mathcal{S}}\int_{\mathcal{Y}_n^{\mathrm{c}}(w_k)\cap\mathcal{Y}_n^{\mathrm{c}}(w_{k+1})}\min\{p(\boldsymbol{y}-f_n(w_k)),p(\boldsymbol{y}-f_n(e_{k+1}))\}\mathrm{d}\boldsymbol{y} \\
=\quad & 2\rho\left(\frac{1}{2M_u},\frac{1}{2M_v}\right)\cdot\frac{1}{M_uM_v}\sum_{k\in\mathcal{S}}\left[\frac{1}{2}\int_{\mathbb{R}^n}\min\{p(\boldsymbol{y}-f_n(w_k)),p(\boldsymbol{y}-f_n(w_{k+1}))\}\mathrm{d}\boldsymbol{y}-\right. \\
& \left.-\frac{1}{2}\int_{\mathcal{Y}_n(w_k)\cup\mathcal{Y}_n(w_{k+1})}\min\{p(\boldsymbol{y}-f_n(w_k)),p(\boldsymbol{y}-f_n(w_{k+1}))\}\mathrm{d}\boldsymbol{y}\right]_+ \\
\overset{(c)}{\geq}\quad & 2\rho\left(\frac{1}{2M_u},\frac{1}{2M_v}\right)\cdot\frac{1}{M_uM_v}\sum_{k\in\mathcal{S}}\left[Q\left(\frac{\|f_n(w_{k+1})-f_n(w_k)\|}{2\sigma}\right)-\right. \\
& \left.\frac{1}{2}\int_{\mathcal{Y}_n(w_k)}p(\boldsymbol{y}-f_n(w_k))\mathrm{d}\boldsymbol{y}-\frac{1}{2}\int_{\mathcal{Y}_n(w_{k+1})}p(\boldsymbol{y}-f_n(w_{k+1}))\mathrm{d}\boldsymbol{y}\right]_+ \\
\overset{(d)}{\geq}\quad & 2\rho\left(\frac{1}{2M_u},\frac{1}{2M_v}\right)\cdot\frac{1}{M_uM_v}\sum_{k\in\mathcal{S}}\left[Q\left(\frac{\|f_n(w_{k+1})-f_n(w_k)\|}{2\sigma}\right)-\frac{\delta_n}{2}-\frac{\delta_n}{2}\right]_+ \\
=\quad & 2\rho\left(\frac{1}{2M_u},\frac{1}{2M_v}\right)\cdot\frac{|\mathcal{S}|}{M_uM_v}\cdot\frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{S}}\left[Q\left(\frac{\|f_n(w_{k+1})-f_n(w_k)\|}{2\sigma}\right)-\delta_n\right]_+
\end{aligned}
$$

$$
\begin{aligned}
&\stackrel{(e)}{\geq} \quad 2\rho\left(\frac{1}{2M_u}, \frac{1}{2M_v}\right) \cdot \frac{(M_u M_v - M_u - M_v)}{M_u M_v} \cdot \left[\frac{1}{|\mathcal{S}|}\sum_{k\in\mathcal{S}} Q\left(\frac{\|f_n(w_{k+1}) - f_n(w_k)\|}{2\sigma}\right) - \delta_n\right]_+ \\
&\stackrel{(f)}{\geq} \quad 2\rho\left(\frac{1}{2M_u}, \frac{1}{2M_v}\right) \cdot \left(1 - \frac{1}{M_u} - \frac{1}{M_v}\right) \times \\
&\qquad \left[Q\left(\frac{1}{2\sigma(M_u M_v - M_u - M_v)}\sum_{k\in S}\|f_n(w_{k+1}) - f_n(w_k)\|\right) - \delta_n\right]_+ \\
&\stackrel{(g)}{\geq} \quad 2\rho\left(\frac{1}{2M_u}, \frac{1}{2M_v}\right) \cdot \left(1 - \frac{1}{M_u} - \frac{1}{M_v}\right) \cdot \left[Q\left(\frac{\sum_i L([f_n]_i)}{2\sigma(M_u M_v - M_u - M_v)}\right) - \delta_n\right]_+, \qquad (25)
\end{aligned}
$$

where $L([f_n]_i)$ is the total length of the part of the signal locus curve, obtained as $w = (u, v)$ travels (continuously) along the i-th diagonal line, $v = (u \cdot M_u + i)/M_v$. This length is well defined thanks to Assumption A.2. The above labeled inequalities are justified as follows:

(a) is by the symmetry of the function ρ and because we are now confining the summation over $\{k\}$ to take place across a subset $\mathcal{S}$ of $\{0, 1, \ldots, M_u M_v - 2\}$, which excludes all $M_u + M_v - 2$ role–over transitions $w_k \to w_{k+1}$, i.e., transitions that are associated with passages from any diagonal line to the next one (in Fig. 1, the transitions $w_0 \to w_1$, $w_2 \to w_3$, $w_5 \to w_6$, $w_9 \to w_{10}$, $w_{12} \to w_{13}$ and $w_{14} \to w_{15}$).

(b) is by the convexity of the function ρ, as $q \geq 1$.

(c) is by the union bound and by recognizing the first integral the line before as the probability of error of the ML decision rule in testing the two equiprobable hypotheses: $\mathcal{H}_0 : \boldsymbol{Y} = f_n(w_k) + \boldsymbol{Z}$ and $\mathcal{H}_1 : \boldsymbol{Y} = f_n(w_{k+1}) + \boldsymbol{Z}$, and by relying on the fact that this probability of error equal to $Q(\|f_n(w_{k+1}) - f_n(w_k)\|/2\sigma)$, where $Q(s) = \frac{1}{\sqrt{2\pi}}\int_s^\infty e^{-t^2/2}\mathrm{d}t$.

(d) is by the outage constraint (8).

(e) is because of the convexity of the function $h(t) = [t - a]_+$ (for any value of a).

(f) is due to the convexity of $Q(s)$ for $s \geq 0$ (which can easily be verified from its second derivative).

(g) is because $Q(\cdot)$ is monotonically decreasing and because $\sum_{k\in S}\|f_n(w_{k+1}) - f_n(w_k)\| \leq \sum_i L([f_n]_i)$, which in turn follows from the fact that the Euclidean distance is a metric, and so, a straight line between two points is never longer than any curve in between.

The last step is to upper bound $\sum_i L([f_n]_i)$ by a quantity that is independent of the particular modulator, f_n, and thereby obtain a universal lower bound to the weak–noise error cost.

Before we proceed, we first provide a brief outline of the parallel step in [13]. In the one–dimensional case of [13], there corresponding quantity was $L(f_n)$ (rather than $\sum_i L([f_n]_i)$), namely, the length of the entire signal locus, and it was upper bounded by a tube–packing consideration: the volume of the object $\{f_n(u) + \boldsymbol{z} :\ u \in [0,1],\ \boldsymbol{z} \in \mathcal{O}_n^{\mathrm{c}}(u),\ \boldsymbol{z} \perp \mathrm{d}f_n(u)\}$ could not exceed the volume of the sphere of radius $\sqrt{n(P+\sigma^2)}$, of typical channel output signals, but on the other hand, it is lower bounded by $L(f_n) \cdot \min_u \mathrm{Vol}\{\mathcal{O}_n^{\mathrm{c}}(u)\}$, which in turn is further lower bounded by $L(f_n)$ times the volume of a sphere of radius (slightly larger than) $\sqrt{n\sigma^2}$, as a sphere occupies the least volume among all objects with a given probability (cf. (8) under the Gaussian density). Thus, $L(f_n)$ was upper bounded in [13] by the ratio between the volumes of spheres with radii $\sqrt{n(P+\sigma^2)}$ and $\sqrt{n\sigma^2}$. This ratio is of the exponential order of $e^{nC(\gamma)}$, similarly as in [15, pp. 672–673] (see also [5, Subsection 2.2.2]).

Here too, we would like to argue that $\sum_i L([f_n]_i)$ is exponentially upper bounded by $e^{nC(\gamma)}$. To this end, we need to be convinced that the spherical tubes surrounding the various parts of the signal locus, $\{[f_n]_i\}$, are disjoint so that the volume of their union would be equal to the sum of the volumes. By part (b) of Definition 1, this in turn will be the case if the parallel diagonal straight lines in the parameter plane are sufficiently far apart. Specifically, for the system to achieve a weak–noise error cost exponent of $E(\gamma) + \epsilon$ (with $\epsilon > 0$ being arbitrarily small), any two points, w and w', along two different parallel diagonal (continuous) straight lines, $v = (u \cdot M_u + i)/M_v$, and $v = (u \cdot M_u + i')/M_v$, must be at distance, $\rho(w - w')$, exponentially no smaller than $e^{-nE(\gamma)}$.

Let us now choose $M_u = M_u^* \triangleq e^{nR_u^*}$ and $M_v = M_v^* \triangleq e^{nR_v^*}$ with

$$R_u^* \triangleq \frac{C(\gamma)}{2} + \frac{a_v - a_u}{2q} \tag{26}$$

$$R_v^* \triangleq \frac{C(\gamma)}{2} + \frac{a_u - a_v}{2q}, \tag{27}$$

which are obtained by implementing eq. (17) for $d = 2$. It can be readily verified by a simple distance calculation, that the ρ–distance between any two points that belong to different diagonal lines, cannot be smaller than $e^{-nE(\gamma)} \min_t[|t|^q + |1-t|^q] = 2^{1-q}e^{-nE(\gamma)}$, which is indeed of the exponential order[5] of $e^{-nE(\gamma)}$. Thus, no matter how small ϵ may be, the tubes surrounding $\{[f_n]_i\}$ must be disjoint (for large enough n), and their total volume cannot exceed the exponential order of $e^{nC(\gamma)}$. But since we chose $M_u M_v = M_u^* M_v^* = e^{n(R_u^* + R_v^*)} = e^{nC(\gamma)}$, the argument of the Q–function in the last line of (25) tends to a constant, and then the lower bound in the last line of (25) becomes of the exponential order of $\rho(e^{-nR_u^*}/2, e^{-nR_v^*}/2) \doteq e^{-nE(\gamma)}$. It follows that $\sup_{F,G,\mathcal{O}} E(F,G,\mathcal{O}) \le E(\gamma)$, which means that a weak–noise error exponent of $E(\gamma) + \epsilon$ cannot be achieved no matter how small

[5] For a general d, the pre-exponential constant factor is different, but the distance between the closest parallel diagonal lines continues to be of the exponential order of $e^{-nE(\gamma)}$.

ϵ is. Since ϵ is arbitrarly small, this completes the proof of Theorem 1.

5 Summary and Outlook

In this paper, we have addressed the problem of modulating a parameter onto a power–limited signal transmitted over a discrete–time Gaussian channel and estimating this parameter at the receiver. The main contribution is a non–trivial extension of the main result of [13], from a single (scalar) parameter (for zero outage exponent) to a multi–dimensional vector parameter. The main conclusions from our results are the following: (i) There is a trade–off among the estimation errors of the various components of the vector parameter. (ii) On a related note, in an optimal modilation–estimation system, all components of the vector parameter contribute essentially evenly to the total weak–noise estimation error cost. (iii) The weak–noise error cost exponent is proportional to the channel capacity and to the power q of the error cost function, and is inversely proportional to the dimension d. (iv) The problem is intimately related to channel capacity theory, both in the direct part and the converse part. Finally, several more general communication system models were briefly discussed.

A few other more general models would be interesting to study in future work. These include the Gaussian colored noise channel model, the presence of clean or noisy feedback, the interference channel, the broadcast channel and more.

References

[1] M. V. Burnashev, "A new lower bound for the a-mean error of parameter transmission over the white Gaussian channel, *IEEE Trans. Inform. Theory*, vol. 30, no. 1, pp. 23–34, January 1984.

[2] M. V. Burnashev, "On minimum attainable mean–square error in transmission of a parameter over a channel with white Gaussian noise, *Problemy Peredachi Informatsii*, vol. 21, no. 4, pp. 3–16, 1985.

[3] D. L. Cohn, *Minimum Mean Square Error Without Coding*, Ph.D. dissertation, Massachusetts Institute of Technology, July 1970.

[4] M. H. M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, no. 5, pp. 439–441, May 1983.

[5] P. A. Floor, *On the Theory of Shannon-Kotelnikov Mappings in Joint Source Channel Coding*, Ph.D. dissertation, Faculty of Information Technology, Mathematics and Electrical Engineering, Department of Electronics and Telecommunications, Norwegian University of Science and Technology, May 2008.

[6] F. Hekland, *On the design and analysis of Shannon-Kotelnikov Mappings for Joint Source Channel Coding*, Ph.D. dissertation, Faculty of Information Technology, Mathematics and Electrical Engineering, Department of Electronics and Telecommunications, Norwegian University of Science and Technology, May 2007.

[7] F. Hekland, P. A. Floor, and T. A. Ramstad, "Shannon-Kotelnikov mappings in joint source-channel coding, *IEEE Trans. Comm.*, vol. 57, no. 1, pp. 94–105, January 2009.

[8] E. Köken, D. Günduz and E. Tuncel, "Energy–distortion exponents in lossy transmission of Gaussian sources over Gaussian channels, *IEEE Trans. on Inform. Theory*, vol. 63, no. 2, pp. 1227–1236, February 2017.

[9] B. Kravtsov and D. Raphaeli, "Multidimensional analog modulation and optimal mappings, *Proc. IEEE Globecom 2007*, pp. 1426–1430, 2007.

[10] N. Merhav, "Universal decoding for memoryless Gaussian channels with a deterministic interference," *IEEE Trans. Inform. Theory*, vol. 39, no. 4, pp. 1261–1269, July 1993.

[11] N. Merhav, "On optimum parameter modulation–estimation from a large deviations perspective," *IEEE Trans. Inform. Theory*, vol. 58, no. 12, pp. 7215–7225, December 2012.

[12] N. Merhav, "Exponential error bounds on parameter modulation–estimation for discrete memoryless channels," *IEEE Trans. Inform. Theory*, vol. 60, no. 2, pp. 832–841, February 2014.

[13] N. Merhav, "Trade-offs between weak–noise estimation performance and putage exponents in nonlinear modulation, to appear in *IEEE Trans. Inform. Theory*, 2019. Also available on–line in `https://arxiv.org/pdf/1802.04627.pdf`

[14] A. Unsal, R. Knopp, and N. Merhav, "Converse bounds on modulation–estimation performance for the Gaussian multiple–access channel," *IEEE Trans. Inform. Theory*, vol. 64, no. 2, pp. 1217–1230, February 2018.

[15] J. M. Wozencraft and I. M. Jacobs, *Principles of Communication Engineering*, John Wiley & Sons Inc., 1965.

[16] R. Zamir, *Lattice Coding for Signals and Networks*, Cambridge University Press, United Kingdom, 2014.