Asymptotic Design of Quantizers for Decentralized MMSE Estimation

Stefano Marano, Vincenzo Matta, Peter Willett, Fellow, IEEE

–DRAFT– update September 19, 2005

Abstract—Conceptual and practical encoding/decoding, aimed at accurately reproducing remotely collected observations, has been heavily investigated since the pioneering works by Shannon about source coding. However, when the goal is not to reproduce the observables, but making inference about an embedded parameter and the scenario consists of many unconnected remote nodes, the landscape is less certain.

We consider a multiterminal system designed for efficiently estimating a random parameter according to the MMSE criterion. The analysis is limited to scalar quantizers followed by a joint entropy encoder, and it is performed in the high-resolution regime where the problem can be easier mathematically tackled.

Focus is made on the peculiarities deriving from the estimation task, as opposed to that of reconstruction, as well as on the multiterminal, as opposite to centralized, character of the inference. The general form of the optimal nonuniform quantizer is derived and examples are given.

I. INTRODUCTION AND BACKGROUND

Consider the scheme of figure 1a, where the observation $x$ is to be recovered at the receiver. In this classical problem, it is well known that a lossless reconstruction of $x$ (assumed discrete) can be achieved by using a rate $R$ not less than the Shannon entropy $H$ of the observable $x$. A lossy reconstruction $\hat{x}$ of a continuous $x$, subject (for instance) to a mean square error constraint $E[(x - \hat{x})^2] \leq D$, can be achieved at a rate $R(D)$ given by the known results about the Shannon rate distortion theory. These issues are dealt with in virtually any textbook about information theory, e.g. [6].

What if the problem at hand is about inference rather than about reconstruction? Consider figure 1b, where the goal is not recovering the observed $x$, but making inference about a parameter $\theta$ that rules the generation of $x$. For instance, if $\theta$ is a random variable, the problem may be mathematically formulated by the joint probability density function (pdf) $f(x|\theta)f_\theta(\theta)$, and one perhaps aims to minimize the mean square error (MSE) $E[(\theta - \hat{\theta})^2]$ at the receiver, with a constraint on the rate $R$. Needless to say, using ex-abrupto known techniques about the data reconstruction may be wasteful, as now the actual matter is still to compress $x$ but with the goal of preserving the information about $\theta$ embodied in $x$.

Following [17], it is instructive to look at the problem as a decentralized (i.e., multiterminal) inference scheme with co-located remote sensors: consider for instance figure 1b’ with two (in general $N$) co-located sensors, so that encoding of $(x_1, x_2)$ can be done jointly. It is well known that the optimal estimator in the MSE sense computed from unquantized observations is the a-posteriori mean $E[\theta|x_1, x_2]$ [19]. So, one can first compute the optimal estimator $\hat{\theta}(x_1, x_2)$ from uncompressed observations $(x_1, x_2)$, and then encode it via standard procedures borrowed from source coding theory [6], [8]. In fact, this approach turns out to be optimal, as shown first by Dobrushin and Tsybakov [7] and subsequently by many other authors, see e.g., [12], [22], [29], [30]. In a lossy scenario, the above topic is closely connected to the classical rate distortion theory, and it is accordingly also referred to as the (moniterminal) indirect rate distortion problem. In fact, let $D_{\text{min}}$ be the minimum mean square error (MMSE) pertaining to the optimal estimator $\hat{\theta}(x_1, x_2)$ built on the unquantized observations, and suppose we wish to encode the data in order to cope with a prescribed MSE distortion $D$. In [7], [12], [22], [29], [30] it is shown that the indirect rate distortion problem can be made equivalent to computing the ordinary...
rate distortion function of the random variable $\hat{\theta}(x_1, x_2)$ at distortion level $D - D_{\min}$, revealing that we can reformulate the whole problem as one of rate distortion, with the role of the source is played by the optimal estimator.

In general, approaching the rate distortion bound requires sophisticated vector quantization techniques. Nonetheless, a practical encoding scheme that is nearly optimum can be obtained by compressing $\hat{\theta}(x_1, x_2)$ using a simple scalar uniform quantizer, followed by an entropy encoder, a well-established technique in the context of quantization for reproduction [10], even for non-high resolution regimes [2], [3], and for correlated sources [34]. According to these studies, the distortion loss due to the scalar nature of the quantizers amounts to $1.53$ dB, a number that we shall again encounter in the following. As the sub-optimality is essentially due to the use of a scalar quantizer, we can loosely say that the above scheme (scalar quantizer followed by entropy encoder) is optimal, within the class of scalar approaches.

Note that quantizing uniformly $\hat{\theta}(x_1, x_2)$ is equivalent to a nonuniform quantization of the observations $(x_1, x_2)$. In other words, there exists an optimal (in the above sense) nonlinear transformation of the data that copes with the inference task in such a way that the transformed data may be successively compressed by classical reproduction-oriented techniques. This is key for the following development: we have recognized that making inference specifically requires the design of a suitable transformation of the observables. After applying such a transformation, a classical compression scheme may be employed. As is intuitive, in the single-terminal case this mapping is just the estimator $\hat{\theta}(x_1, x_2)$.

Let us turn to the multiterminal scenario. With reference to figure 1c, two (in general $N$) remote nodes observe $(x_1, x_2)$ drawn from the a known joint density $f(x_1, x_2)$, and $i^{th}$ node has to encode its observation $x_i$, separately from other nodes (nodes cannot communicate). If perfect (lossless) reconstruction is desired, a milestone theorem by Slepian and Wolf (SW) tells us that, assuming $R_1 = R_2$ for symmetry, this can be achieved with a sum rate equals the joint entropy of the observations $(x_1, x_2)$: the same efficiency of as with co-located quantizers is achieved [23]. Less general results are known for lossy source coding, although there has been considerable study.

As for the single-terminal examples of before, we now switch to the inference (as opposed to reconstruction) problem, see figure 1d. The issue can be traced back to 1979 when Berger [1] formulated a problem that, following [15], “was the first attempt to combine two seemingly different kinds of problems which had been separately investigated in the void between statistics and information theory.” The setup of [1], see also [15], [33], was that of nonrandom parameter estimation. Subsequently, Yamamoto [31], Flynn and Gray [13], and, more recently, Berger [4], adopted the different setting where the unknown $\theta$ is modelled as a random variable. The present paper refers to this latter scenario.

Let us now briefly summarize some key results from the literature. Gish and Pierce [10] tell us that uniform quantizers followed by entropy encoders are nearly optimal in a single-terminal reproduction-oriented scenario, a (high-rate) result corroborated in [14] even in the vector quantization (VQ) case: quantization points form a lattice and even were collaborative quantization among the sources possible the optimal scheme would not use it. Placing this insight in a multiterminal setting, Zamir and Berger [32] show that uniform (actually lattice) quantizers followed by SW encoders are nearly optimal for reproduction purposes, in the high resolution regime.

The main question we address in this paper immediately arises:

Assuming we use SW encoders, how should the quantizers be designed for inference purposes in a multiterminal setting?

To answer, we capitalize on an important paper by Poor [20] dealing with the $f$-divergence performance loss due to high resolution data quantization. As many of the commonly adopted inference performance figures are special forms of the $f$-divergence, it comes with no surprise that some of the relationships presented in this paper can be seen as special case of Poor’s general formulas. Also key to our development is the already quoted paper by Gish and Pierce [10]. Note that this latter work neither considers inference nor multiterminal situations, while in [20] the focus is on inference, but no SW encoding is assumed, implying that a different optimization problem is dealt with.

Let us finally recall that there exists a considerable literature addressing the topic of finding the optimal rate distortion region for multiterminal estimation. A complete solution is known only for a few special cases [4], [5], [18], [24], [26], and in the following we will make some reference to one of this: the quadratic Gaussian CEO problem [5].

The remainder of this paper is organized as follows. The next section introduces the system architecture and puts forth some premises. Section III deals with the evaluation of rate and distortion figures, and with the system optimization. Section IV gives the general formulas for an arbitrary number of sensors, compares the rate distortion curves of the designed optimized system to those obtained with uniform quantizers, and provides some asymptotic relationships. Numerical examples are offered in Section V, while we summarize in Section VI. Some mathematical details are postponed to an Appendix.

II. DISTRIBUTED SYSTEM ARCHITECTURE

Decentralized inference cannot be handled by simple generalization of the single-terminal counterpart, except for the trivial case earlier discussed of co-located sensors. There is no hope of computing the optimal MMSE estimator $\hat{\theta}(x_1, x_2)$ at any individual remote sensor and, on the other hand, the optimal estimator $\hat{\theta}(x_1, x_2)$ usually does not decouple into a combination of the optimal local estimates, say $\hat{\theta}(x_1)$ and $\hat{\theta}(x_2)$.

1To avoid confusion we stress that the scalar/vector nomenclature refers to the compression of single/multiple observations at any given one among the sensors, not the fact that the system is mono/multi-terminal.

2Actually, [32] also addresses the case of inference, but under Gaussian assumptions.
The goal of this paper is to design optimal (but scalar) quantizers specifically tailored to inference purposes, thus trying to establish useful design guidelines that generalize those of the single-terminal case. First, let us observe that there may exist cases in which a remote node can form a local sufficient statistics $T_i(x_i)$. In these instances, it would make more sense to quantize these local statistics rather than the raw data $x_i$’s [11], and the following theory is to be applied with $x_i$ replaced by $T_i(x_i)$, $\forall i$. For simplicity, we do not address this scenario\(^3\).

The system architecture, referring to the case of figure 1d (for most part of this work we shall deal with two sensors), is sketched in figure 2. A common source of information is modeled as a random variable $\theta$ of known pdf $f_\theta(\theta)$. The remote unconnected sensors observe realizations of two continuous random variables $x_1$ and $x_2$ with joint density$^4$ $f(x_1, x_2|\theta)f_\theta(\theta)$. A case of practical relevance is of conditionally independent observations, in which $f(x_1, x_2|\theta) = f(x_1|\theta)f(x_2|\theta)$, but if not otherwise stated we address the more general scenario. We also assume that the variables $x_i$ are exchangeable, i.e., $f(x_1, x_2|\theta)$ is symmetric in $(x_1, x_2)$, thus complying with a physical model of identical sensors operating in the same environment.

At the encoding stage, the observations $x_1$ and $x_2$ collected at the sensors are processed by means of scalar quantizers, producing outputs $q_1$ and $q_2$. These discrete quantities are then compressed via Slepian-Wolf encoders, which are assumed perfectly lossless and achieving a sum rate $R = R_1 + R_2$. This latter, complying with the SW theorem, is given by the joint entropy of the random variables $q_1$ and $q_2$. At the decoding stage, the received data are first jointly decoded by the SW-decoder, and then fused to the MMSE estimate of $\theta$ from the quantized observations $q_1$ and $q_2$:

$$\widehat{\theta}_q = \int \theta f_\theta(\theta|q_1, q_2)d\theta,$$

where $f_\theta(\theta|q_1, q_2)$ being the a-posteriori density of the random variable $\theta$, given the observables $q_1$ and $q_2$.

The system architecture described in figure 2 follows from the model considered by Zamir and Berger in [32]. Paralleling the earlier discussion of the single-terminal case, the rationale is that the SW coding (optimally) deals with the reconstruction stage, while the quantizer design, amounting to finding a suitable data transformation, copes with the inference task. Assuming a perfect SW encoding/decoding schemes (currently for practical systems only a goal), we are leaving aside any reconstruction concern to focus on the inference issue.

Summarizing, we limit the scope to the scalar case and our goal is to design the quantizers in the context of inference. The original (symmetric, exchangeable) observations $x_i$ should be maximally compressed, not with the goal of preserving the quality of the recovered $x_1$ and $x_2$, but instead of maximizing the quality of the inference about $\theta$. We address this in the high-rate regime where, as shown below, the problem is mathematically tractable. Numerical work shows that conclusions apply in non-asymptotic cases as well.

### III. Quantizer Design

The following derivation parallels the optimization proposed in the classical paper by Gish and Pierce [10] for the case of reproduction.

We denote the sensors’ quantized outputs by

$$q_1 = Q(x_1), \quad q_2 = Q(x_2),$$

where the quantizer input/output function $Q$, for simplicity the same for all sensors, is uniquely determined by the partition regions $\mathbb{R}_n$ and by the restitution levels $r_n$:

$$Q(x) = r_n \quad \text{if } x \in \mathbb{R}_n,$$

meaning that $q_i, i = 1, 2,$ may take one of the possible values $\{r_n\}$. The partition regions can be defined in terms of a suitable smooth, monotone increasing function $g(x)$: we have

$$\mathbb{R}_n = (g_n, g_{n+1}),$$

where $g_n = g(n\delta)$, and $\delta$ is the a parameter ruling the quantizer resolution. Otherwise stated, as illustrated in the zoom of figure 2, a general (i.e., nonuniform scalar) quantizer can be always obtained using a uniform quantizer and an appropriate nonlinearity $g(x)$ [8]. For our inference purposes the restitution levels are immaterial, it is sufficient to know at which region $\mathbb{R}_n$ a given $r_n$ belongs to. This means that the restitution levels of the uniform quantizer depicted in figure 2 may be left unspecified. As $Q(x)$ is defined in terms of $g(x)$, the quantizer design amounts to appropriately selecting this latter nonlinearity, and this is the goal of our optimization procedure.

\(^3\)For example, if the optimal estimator is known to be a function of the squared observations $x_i^2$, then compressing $|x_i|$ would be better than compressing $x_i$; preserving the sign is wasteful.

\(^4\)To simplify the notation we occasionally denote with $f$ many different probability density functions modeling the unquantized observations $x_i$; the number of arguments, the possible presence of conditioning, and the context should avoid confusions. When the pdf refers to other quantities, for instance to $\theta$, an explicit subindex will be appended.
A. Rate Requirements

Let us now compute the rate requirements of the decentralized system shown in figure 2. As said, employing SW encoders allows to identify the (sum-)rate as the joint entropy of the quantizer’s outputs, that is

$$R = H = - \sum \sum p_{mn} \log p_{mn},$$  \hspace{1cm} (4)

where \(p_{mn}\) is the joint probability mass function (pmf) of the quantized observations

\[ p_{mn} = \Pr(x_1 \in \mathcal{R}_m, \ x_2 \in \mathcal{R}_n). \]

In the high-resolution regime \(\delta \to 0\), the above pmf can be approximated as

\[ p_{mn} \approx f(g_m, g_n)(g_{m+1} - g_m)(g_{n+1} - g_n), \]

where \(f(x_1, x_2)\) is the joint pdf of the unquantized data \(x_1, x_2\). On the other hand, for small \(\delta\) it also holds that \(g_{m+1} - g_m \approx \delta g(m\delta)\), where the dot denotes the derivative with respect to the argument. Thus eq. (4) can be rewritten as:

\[ R \approx - \int \int f(t, \tau) \hat{g}(t) \hat{g}(\tau) \log \left[ f(t, \tau) \delta^2 \hat{g}(m\delta) \hat{g}(n\delta) \right] dt d\tau. \hspace{1cm} (6) \]

By the substitution \(\{g(t), g(\tau)\} \to \{s, \phi\}\) and further defining

\[ \gamma(s) = \hat{g}\left(g^{-1}(s)\right), \]

one finally gets

\[ R \approx - \int \int f(s, \phi) \log \left[ f(s, \phi) \delta^2 \gamma(s) \gamma(\phi) \right] ds d\phi. \hspace{1cm} (8) \]

The rate is now easily represented by splitting the integral in (8) as:

\[ R \approx h(x_1, x_2) - 2 \log \delta - 2 \int f(s) \log \gamma(s) ds \hspace{1cm} (9) \]

where \(h(x_1, x_2)\) is the joint differential entropy of \(x_1\) and \(x_2\).

As checks of the above expression, note that if \(Q\) is uniform, namely if \(\gamma(s) = 1\), eq. (9) yields \(R = h(x_1, x_2) - 2 \log \delta\), which is a simple generalization of the well known result for a single random variable [6]. Further, if a centralized system with a single sensor were considered, the result of [10] arises. Generalization can be obtained with the tools provided in [20].

\[ ^{\text{Unless otherwise specified, we adopt the logarithm to base 2, and denote it by log. The entropies are accordingly measured in bits. We reiterate that in all our development we for simplicity use only two sensors. The same arguments and conclusions apply for the general case.}} \]

B. Distortion Evaluation

Let us come now to the computation of the distortion. We recall that the MMSE estimators stemming from unquantized and quantized data are, respectively:

\[ \hat{\theta}_x(x_1, x_2) = E[\theta|x_1, x_2], \quad \hat{\theta}_q(q_1, q_2) = E[\theta|q_1, q_2]. \]

The selected distortion \(D\) measures the mean square distance between the optimal MMSE estimator computed from quantized data \(\hat{\theta}_q(q_1, q_2)\), and the parameter to be estimated \(\theta\):

\[ D = E \left[ (\hat{\theta}_x - \theta)^2 \right]. \]

Adding and subtracting \(\hat{\theta}_x(x_1, x_2)\), and omitting the dependence upon \((x_1, x_2)\) and \((q_1, q_2)\) for notational convenience, yields

\[ D = E \left[ (\hat{\theta}_x - \theta)^2 \right] + E \left[ (\hat{\theta}_x - \hat{\theta}_q)^2 \right] - 2E \left[ (\hat{\theta}_x - \theta)(\hat{\theta}_x - \hat{\theta}_q) \right]. \hspace{1cm} (10) \]

It is well known that, in view of the orthogonality principle, the error for MMSE estimates is orthogonal to the data, as well as to any measurable function thereof [19]. This implies that the last addend in eq. (10) is zero. Thus, we have

\[ D = D_{\text{min}} + E[\hat{\theta}_x(x_1, x_2) - \hat{\theta}_q(q_1, q_2)] \]

where \(D_{\text{min}} = E[(\hat{\theta}_x(x_1, x_2) - \theta)^2]\) is the minimum achievable distortion if the unquantized version of the data were fully available at the receiving stage. Clearly, the quantizer design has no effect on \(D_{\text{min}}\) so that minimizing the second addend in eq. (11) is the actual goal.

In the high resolution regime the partition regions are sufficiently small that a linear approximation for \(\hat{\theta}_x(x_1, x_2)\) is appropriate. Since \(\hat{\theta}_q(q_1, q_2) = E[\hat{\theta}_x(x_1, x_2)|q_1, q_2]\), we can exchange the average operator with the (assumed linear) function \(\theta_x(x_1, x_2)\), thus obtaining

\[ \hat{\theta}_q(q_1, q_2) \approx \hat{\theta}_x(v_1, v_2), \]

where \(v_1\) and \(v_2\) are the centroids of the partition cell given the pair \((q_1, q_2)\). Equation (11) can be accordingly recast as

\[ D = D_{\text{min}} + E \left[ (\hat{\theta}_x(x_1, x_2) - \hat{\theta}_x(v_1, v_2))^2 \right]. \hspace{1cm} (12) \]

Capitalizing on the high-rate assumption, we have

\[ \hat{\theta}_x(x_1, x_2) - \hat{\theta}_x(v_1, v_2) \approx \frac{\partial \hat{\theta}_x(x_1, x_2)}{\partial x_1} (x_1 - v_1) + \frac{\partial \hat{\theta}_x(x_1, x_2)}{\partial x_2} (x_2 - v_2). \hspace{1cm} (13) \]

In the high-resolution regime, the centroids \(v_1\) and \(v_2\) can be exchanged with the geometric centroids [27], that is, in our case, with the midpoints of the partition regions. Furthermore, the double product term arising from substituting eq. (13) in the statistical average of eq. (12) vanishes, in view of the asymptotic whiteness\(^7\) of the quantization errors [27]. The two

\[ ^{\text{We stress that the restitution levels (centroids) are computed at the receiver stage: encoders are separate but decoding is performed jointly.}} \]

\[ ^{\text{The presence of the partial derivatives in eq. (13) does not invalid these results, provided that the derivatives are sufficiently smooth.}} \]
surviving squared terms are identical for symmetry reasons, so that the statistical average in eq. (12) reduces to
\[ 2E \left[ \left( \frac{\partial H(x_1, x_2)}{\partial x_1} \right)^2 (x_1 - \nu_1)^2 \right]. \quad (14) \]

It is now convenient to make explicit the expectation in (14) in the form
\[ \int \left[ \int \left( \frac{\partial H(x_1, x_2)}{\partial x_1} \right)^2 f(x_2|x_1) dx_2 \right] (x_1 - \nu_1)^2 f(x_1) dx_1. \]

Letting
\[ w^2(x_1) = \int \left( \frac{\partial H(x_1, x_2)}{\partial x_1} \right)^2 f(x_2|x_1) dx_2 \quad (15) \]
the distortion becomes
\[ D = D_{\min} + 2 \int w^2(x_1)(x_1 - \nu_1)^2 f(x_1) dx_1. \quad (16) \]

The whole distortion is a sum of two terms, the first of which, \( D_{\min} \), is the lower bound that one could achieve only if no quantization at all would be involved. The interesting term is the second in which the factor 2 arises from our initial assumption of considering two identical sensors. Let us consider the physical meaning of this distortion term from the perspective of sensor #1. Equation (16) tells us that sensor #1 contributes to the inference distortion with a term which is a sort of reproduction error \( (x_1 - \nu_1)^2 \) weighted by a function \( w^2(x_1) \) that accounts for the fact that we are dealing with an inference (estimation) problem. In fact, as seen from eq. (15), \( w^2(x_1) \) strongly depends upon the partial derivative of the optimal estimator from unquantized observations in the direction of \( x_1 \). Thus, this derivative represents the influence that an error in reconstructing \( x_1 \) implies on the final estimation error. Note also that, not unexpectedly, the partial derivative is averaged over the unseen measurement \( x_2 \) given the observed \( x_1 \), as perfectly sounding from the point of view of sensor #1.

Let us now come back to the computation of the distortion for optimization purposes. The integral (16) can be split over the partition regions:
\[ D = D_{\min} + 2 \sum \int_{\mathbb{R}_n} w^2(x)(x - r_n)^2 f(x) dx, \quad (17) \]
where \( r_n \) is the midpoint of \( \mathbb{R}_n = (g_n, g_{n+1}) \), see discussion following eq. (13). Further, \( \mathbb{R}_n, (17) \) can be approximated by
\[ D \approx D_{\min} + 2 \sum w^2(g_n)f(g_n)(g_{n+1} - g_n)^3/12 \approx D_{\min} + 2 \sum w^2(g_n)f(g_n)(\delta g(n\delta))^3/12, \]
with the sum approaching an integral:
\[ D \approx D_{\min} + 2/12 \int f(g(t)) [w(g(t)) \delta g(t)]^2 \delta g(t) dt. \quad (18) \]
The substitution \( g(t) \to s \) finally gives
\[ D \approx D_{\min} + 2(\delta^2/12) \int f(s) [w(s)g(s)]^2 ds. \quad (19) \]
Formula (19) is a special case of a more general result in [20].

C. Rate Distortion Optimization

The main results are summarized by eqs. (9) and (19):
\[ R \approx h(x_1, x_2) - 2\log \delta - 2 \int f(s) \log g(s) ds, \]
\[ D \approx D_{\min} + 2(\delta^2/12) \int f(s)[w(s)g(s)]^2 ds. \]

Using the two above relationships, we want to find the best nonlinearity \( g(\cdot) \) that, for any prescribed \( \delta \) and \( D \), minimizes the rate \( R \). To this aim, we let \( \Gamma(s) = w(s)g(s) \), thus obtaining
\[ R \approx h(x_1, x_2) - 2\log \delta + 2 \int f(s) \log w(s) ds - 2 \int f(s) \log \Gamma(s) ds, \]
\[ D \approx D_{\min} + 2(\delta^2/12) \int f(s)\Gamma^2(s) ds. \]

The solution to the optimization problem is now equivalent to that discussed in [10]: for any \( \delta \) and \( D \), \( R \) is minimum when the function \( \Gamma(s) \) is a constant. This constant can be safely taken as unity\(^8\), yielding \( \gamma(s) = 1/w(s) \). From definition (7) we have
\[ \gamma(s) = \dot{g}(g^{-1}(s)) = 1/w(s) \Leftrightarrow \frac{d}{ds} g^{-1}(s) = w(s). \quad (22) \]
and the nonlinearity sought is found in integral form:
\[ g^{-1}(s) = \int_a^s w(\phi) d\phi \quad (23) \]
where a typical choice of the constant \( a \) is such that \( g^{-1}(0) = 0 \), for symmetry reasons, see [8].

The above equation allows us to design the optimal quantizers, provided that the MMSE estimator \( \theta_x(x_1, x_2) \) is found. The method is illustrated in the section devoted to the examples.

IV. Rate Distortion Curves and Arbitrary \( N \)

The above derivation considers two sensors, and in principle it can be extended to an arbitrary number \( N \) of nodes, yielding
\[ R \approx h - N \log \delta + N \int f(s) \log w(s) ds \]
\[ - N \int f(s) \log \Gamma(s) ds, \]
\[ D \approx D_{\min} + N(\delta^2/12) \int f(s)\Gamma^2(s) ds. \]

Note however that deriving out the optimum nonlinearity for large \( N \) is difficult, in that eq. (15) becomes a \((N - 1)\)-dimensional integral:
\[ w^2(x_1) = \int \left( \frac{\partial H(x_1, \ldots, x_N)}{\partial x_1} \right)^2 f(x_2, \ldots, x_N|x_1) dx_2, \ldots, dx_N. \quad (26) \]
Solving this usually requires Monte Carlo integration [9].

\(^8\)This indeterminacy can be explained. Consider a solution \( \Gamma^*(s) = c\delta(s) \). This is equivalent to scaling in both the eqs. (20) and (21) the resolution \( \delta \) to \( c\delta \). This simply stretches the \( \delta \)-axis, thus exploring a different point of the operational rate distortion curve.
An alternative approach is based on optimizing a lower bound of the MSE, rather than the exact distortion $D$. We thus refer to the so called Van Trees (or Bayesian Cramér-Rao) inequality [25], that lower bounds the MSE for the estimation of random parameters, see eq. (41) in the Appendix. This performance figure has been already profitably adopted in the specific context of decentralized estimation systems (see, e.g., the converse theorem for the quadratic Gaussian CEO problem [26]). The technique we propose requires conditional independence of the $x_i$’s given $\theta$ (although inequality (41) is valid under far more general settings) and we hence assume that.

Adopting the Van Trees bound as distortion metric and paralleling the previous approach, in the Appendix we show that expressions identical to eqs. (24) and (25) can be derived. The weighting function defining the quantizers’ nonlinearity becomes, see eq. (47):

$$w^2(x) = \frac{1}{(I_\theta + N I_x)^2} \left( \frac{\partial^2 \ln f(x|\theta)}{\partial x \partial \theta} \right)^2 f_\theta(x|\theta) d\theta, \quad (27)$$

where the quantities $I_\theta$ and $I_x$ are defined in the Appendix. It is immediately recognized that the value of $N$ makes no difference at all, and we thus circumvent our concern of eq. (26).

In the following subsections we combine eqs. (24) and (25) to get the operational rate distortion curve. This latter is first specialized to the designed optimized system and then, for comparison purposes, to an unoptimized system employing simple uniform quantizers. In the section devoted to the examples, the operational rate distortion curves (optimal system and uniform quantizers) are illustrated for specific examples. There, whenever appropriate, the approach based on the Van Trees bound is also exploited.

A. Optimized System

For the optimized system (i.e., $\Gamma(s) = 1$),

$$R \approx h - N \log \delta + N \int f(s) \log w(s) ds$$

$$D \approx D_{\text{min}} + N \delta^2 / 12,$$

Let us set

$$\kappa(N) = 2 \int f(s) \log w(s) ds, \quad (28)$$

where the dependence upon the number of sensors is embodied in the weighting function $w(s)$, see eq. (26). The previous expressions of $R$ and $D$ become

$$R \approx h - N \log \left( \frac{\delta}{\sqrt{\kappa(N)}} \right);$$

$$D \approx D_{\text{min}} + N \delta^2 / 12,$$

that, solving for $\delta$, provide us with the final rate distortion curves for the optimized system:

$$R(D) = h - \frac{N}{2} \left[ \log \left( \frac{12}{N \kappa(N)} (D - D_{\text{min}}) \right) \right]. \quad (29)$$

B. Uniform Quantization

For uniform quantizers (i.e., $\gamma(s) = 1$, yielding $\Gamma(s) = w(s)$), we get

$$R_u \approx h - N \log \delta,$$

$$D_u \approx D_{\text{min}} + N \kappa_u(N) \delta^2 / 12,$$

where we have set

$$\kappa_u(N) = \int f(s) w^2(s) ds.$$

Combining the above in a single equation, we have

$$R_u(D_u) = h - \frac{N}{2} \left[ \log \left( \frac{12}{N \kappa_u(N)} (D_u - D_{\text{min}}) \right) \right]. \quad (30)$$

In the absence of specific guidelines for optimal quantizer design, one would certainly be tempted to employ uniform quantization. As the quantizer stage is followed by a SW (i.e., entropy) encoder, and having in mind the lesson learned in the monoterinal scenario (the Gish/Pierce result), one might hope that the loss be small. Actually, as shown in this paper, in general it is not so, and uniform quantization thus represents a meaningful benchmark for testing the theory we are developing.

C. Optimized vs Uniform

We can now easily compare the operational curves of the optimized system against the blind compression scheme employing uniform quantizers. For a prescribed distortion $D^*$, we get, from eqs. (29) and (30):

$$\Delta R = R_u(D^*) - R(D^*) = \frac{N}{2} \log \left( \frac{\kappa_u(N)}{\kappa(N)} \right) = \frac{N \zeta(N)}{2}, \quad (31)$$

where $\zeta(N) = \log(\kappa_u(N) / \kappa(N))$. We might equivalently evaluate the two distortions for a given rate $R^*$. It makes more sense to make a comparison between the two systems in terms of the error $D = D - D_{\text{min}}$: we have

$$\Delta D_{\text{DB}} = 10 \log_{10} \frac{D_u(R^*) - D_{\text{min}}}{D(R^*) - D_{\text{min}}}$$

$$= 3.01 \log \left( \frac{\kappa_u(N)}{\kappa(N)} \right) = 3.01 \zeta(N). \quad (32)$$

V. EXAMPLES

The physical problem is entirely specified in terms of $f_\theta(\theta)$ and $f(x|\theta)$, where $f_\theta(\theta)$ is the prior distribution of the parameter, and $f(x|\theta)$ represents the joint conditional distribution of the random variables $x = [x_1, x_2, \ldots, x_N]$ pertaining to the $N$ sensors. Three out of the four provided examples assume that $f(x|\theta) = \prod_{i=1}^N f(x_i|\theta)$, meaning that the sensors are conditionally independent given the parameter $\theta$.

In order to tune the degree of nonlinearity of the distributed estimation problem in a simple way, in some cases we find it convenient to introduce an auxiliary set of random variables $y = [y_1, y_2, \ldots, y_N]$, and assume that the observed $x_i$’s are a one-to-one nonlinear transformation thereof $x_i = t^{-1}(y_i)$, $\forall i$. In this way, we can first assign to the $y_i$’s a distribution
\(f_y(y|\theta)\) in a convenient analytical form, and then consider an appropriate transformation \(t(\cdot)\) ruling the \(x_i\)'s statistics and imposing the desired degree of nonlinearity. Note that, obviously, should the \(x_i\)'s be conditionally independent, the same property would apply to the transformed \(y_i\). Further, if the MMSE estimator \(\hat{y}_N(y_1, y_2, \ldots, y_N)\) stemming from the \(y\)-data is known, then the MMSE estimator built on the actually observations \(x_i\)'s is related to that by

\[
\hat{\theta}_x(x_1, x_2, \ldots, x_N) = \hat{\theta}_y(t(x_1), t(x_2), \ldots, t(x_N)). \tag{33}
\]

### A. Gaussian Inference and the CEO Problem

Given \(\theta\), the vector \(y\) is Gaussian and the transformation \(t(\cdot)\) is the identity: \(x = y\). Specifically, we assume

\[
f(x_i|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{ -\frac{(x_i - \theta)^2}{2\sigma^2} \right\} = N(x_i; \theta, \sigma),
\]

with a Gaussian prior \(f_0(\theta) = N(\theta; 0, \sigma_\theta)\). This is usually referred to as the Gaussian CEO problem [4]. The MMSE estimator of \(\theta\) is found in the form

\[
\hat{\theta}_x(x_1, x_2, \ldots, x_N) = \frac{\sigma^2}{\sigma^2 + N\sigma^2_\theta} \sum_{i=1}^{N} x_i,
\]

and its partial derivative in the direction \(x_i\) is

\[
\frac{\partial \hat{\theta}_x(x_1, x_2, \ldots, x_N)}{\partial x_i} = \frac{\sigma^2}{\sigma^2 + N\sigma^2_\theta},
\]

so that the weighting function defined in eq. (15) amounts to

\[
w(s) = \frac{\sigma^2}{\sigma^2 + N\sigma^2_\theta}. \tag{34}
\]

This, in view of eq. (23), reveals that the optimal nonlinearity is a straight line. The above results hold for an arbitrary number of sensors \(N\). In the case at hand, computing the nonlinearity following the approach based on Van Trees bound is a simple task and, not unexpectedly, this yields the same conclusion: \(g(x)\) must be linear.

Figure 3 shows the shape of the optimum nonlinearities for the offered examples. The Gaussian case at hand, depicted for comparison with the following examples, is reported in the top-left panel: for all \(N\) the optimal compandor curve is actually a straight line.

Now, recall that the optimum nonlinearity has been specifically designed for exploiting the distinct feature of the inference problem and that a uniform quantization would be the best choice if reproduction were addressed. In this respect, and perhaps unsurprisingly, the inference problem for the Gaussian CEO scenario seems to degenerate to a reproduction one, with the (SW-coded) uniform quantizer being optimal. Conceptually, the optimum estimator is built by first recovering the whole data \(x_i\), and then exploiting these reconstructed quantities for the final \(\theta\)-estimate.

For the Gaussian CEO, the optimal rate distortion curves are known, and this allows to perform a sanity check by comparing the performances of the designed system with these theoretical bounds. Since our results were derived in the high-resolution regime, we need to address the comparison under such assumption. Then, the pertinent rate distortion curve for the Gaussian CEO problem is provided in [32]:

\[
R_{CEO}(D_{CEO}) = h - \frac{N}{2} \log \left(2\pi e N(D_{CEO} - D_{min})\right). \tag{35}
\]

This must be compared to eq. (29) representing our optimized system. The quantity \(\kappa(N)\) defined by eq. (28) can be computed via the expression of \(w(s)\) in eq. (34), yielding

\[
\kappa(N) = \left(\frac{\sigma^2}{\sigma^2 + N\sigma^2_\theta}\right)^2. \tag{36}
\]

For large \(N\), we get \(\kappa(N) \approx N^{-2}\), and eq. (29) reduces to

\[
R(D) = h - \frac{N}{2} \log (12N(D - D_{min})). \tag{37}
\]

Let us prescribe a certain rate \(R^*\): we have

\[
10 \log_{10} \frac{D(R^*) - D_{min}}{D_{CEO}(R^*) - D_{min}} = (10 \log_{10} 2) \log \frac{\pi e}{6} = 1.53 \text{ dB}.
\]

This is the classic 1.53 dB of loss that is incurred, in the context of quantization for reconstruction purposes, as consequence of employing entropy-coded uniform scalar quantizers [10]. We conclude that the designed system is optimal, given the a-priori choice of quantizing in a scalar manner the observations \(x_i\). If remote nodes would employ more complicated vector quantizers we expect that the loss would accordingly reduce.
Now, however, the mapping \( x \) runs. Note how the design of the optimum nonlinearity (NL) substantially outperforms the system exploiting entropy encoded uniform quantizers.

### B. Transformed Gaussian Case

As before, the vector \( y \) is conditionally Gaussian \( f_Y(y_i|\theta) = \mathcal{N}(y_i; \theta, \sigma) \), as is the prior \( f_\theta(\theta) = \mathcal{N}(\theta; 0, \sigma_\theta) \). Now, however, the mapping \( x_i = t(y_i) \) is the well-known \( \mu \)-law nonlinearity [8]; that is, an odd function defined, for \( x \geq 0 \), by

\[
t(x) = \frac{\ln(1 + x\mu/V)}{\ln(1 + \mu)} V, \tag{36}
\]

where \( \mu \) is a parameter ruling the degree of nonlinearity and \( V \) is such that \( t(V) = V \). In terms of the fictitious \( y \)-data, the optimal MMSE estimator is as in eq. (34):

\[
\hat{\theta}_y(y_1, y_2, \ldots, y_N) = \frac{\sigma_\theta^2}{\sigma^2 + N\sigma_\theta^2} \sum_{i=1}^{N} y_i,
\]

and in view of eq. (33) this implies

\[
\hat{\theta}_x(x_1, x_2, \ldots, x_N) = \frac{\sigma_\theta^2}{\sigma^2 + N\sigma_\theta^2} \sum_{i=1}^{N} t(x_i). \tag{37}
\]

The additive form of such estimator (see also [21]) allows us to evaluate very easily the relevant nonlinearity. In fact, the partial derivative of the optimal estimator is

\[
\frac{\partial \hat{\theta}_x(x_1, x_2, \ldots, x_N)}{\partial x_i} = \frac{\sigma_\theta^2}{\sigma^2 + N\sigma_\theta^2} t(x_i), \tag{38}
\]

whence the optimal nonlinearity results

\[
g^{-1}(s) = \frac{\sigma_\theta^2}{\sigma^2 + N\sigma_\theta^2} t(s), \tag{39}
\]

having enforced the condition \( g^{-1}(0) = 0 \), complying with the inherent symmetry of the problem. This amounts to saying that, for such additive estimators, the optimum nonlinearity is, but for an irrelevant multiplying factor, independent of the number of sensors, and only depends upon the transformation \( t(x) \).

The top-right panel of Fig. 3, which refers to the present example, shows that the inference-oriented quantizer may be highly nonuniform. As said, and as happened in the previous example, since the optimum nonlinearity does not depend upon \( N \), the quantizers are one and the same as that used in a single-terminal system.

Let us now evaluate the optimal nonlinearity via Van Trees bound. We have:

\[
\frac{\partial \ln f(s|\theta)}{\partial \theta} = \frac{\partial \ln (f_Y(t(s)|\theta) t(s))}{\partial \theta} = t(s) - \theta \frac{\sigma^2}{\sigma^2}, \tag{40}
\]

where the last equality follows from the Gaussianity of the \( y_i \)'s.

Furthermore, it is relatively easy to compute \( I_x = 1/\sigma^2 \) and \( I_\theta = 1/\sigma_\theta^2 \) (see the Appendix for the definitions of \( I_x \) and \( I_\theta \)), so that eq. (27) gives

\[
w(s) = \frac{\sigma_\theta^2}{\sigma^2 + N\sigma_\theta^2} t(s).
\]

This immediately yields the same \( g^{-1}(s) \) as in eq. (39).

As the optimal quantizer is not uniform, it is interesting to compare the rate distortion performances of the optimized system to those of a system employing uniform quantizers. Figure 4 depicts \( D - D_{\text{min}} \) versus the rate \( R \), as described by eqs. (29) (optimized system) and (30) (uniform quantizers) for a decentralized system made of two sensors. The relevant quantities \( \kappa(N)|_{\eta=2} \) and \( \kappa_u(N)|_{\eta=2} \) have been computed numerically and they determine, through \( \zeta(N) \), the gain of the optimized system over that employing uniform quantizers: we have \( \zeta(N) = 2.82 \) which amounts to \( \Delta D_{\text{AB}} = 8.5 \). For the case at hand \( \zeta(N) \) actually does not depend upon \( N \), as can be easily checked from eq. (38). Perhaps, this comes with no surprise as in many cases the behavior of \( \kappa(N) \) and \( \kappa_u(N) \)
with respect to $N$ is the same; this happens, for instance, when the estimator $\hat{\theta}_e$ essentially depends upon $N$ via a multiplying constant as in eq. (37).

As expected, the points obtained by Monte Carlo simulation progressively approach the theoretical curves, as the resolution increases. It is also worth noting that, at low rates (where the theoretical curve loses validity since it is a high-rate result), the two systems tend to behave similarly, and this may be explained by considering the limiting case of one-bit quantization. Then, in fact, the uniform quantizer simply discerns the observation’s sign, and the same happens with our optimized system, the optimum nonlinearity being an odd function. Figure 5 shows the rate distortion curve for $N = 10$. Comments similar to the $N = 2$ case apply: (i) the distortion gain is the same as before, as seen in eq. (32), and (ii) the rate gain grows proportionally to $N$, see eq. (31). The rates plotted are the sum rates, and it is observed that the per-sensor rate necessary for a given distortion is lower when there are more sensors.

C. Gaussian Mixture Case

Let us now consider the case that, given $\theta$, each $y_i$ is a balanced mixture of two Gaussian variables

$$f_Y(y_i|\theta) = \frac{1}{2}N(y_i; \theta_1, \sigma_1) + \frac{1}{2}N(y_i; \theta_2, \sigma_2),$$

with (again) a Gaussian prior $f_\theta(\theta) = N(\theta; 0, \sigma_\theta)$. The $x_i$’s result from $x_i = t^{-1}(y_i)$ and the mapping is again given by eq. (36). The optimal estimator$^9$ is not of the simple form previously found, namely it does not decouple. As a consequence, it is now difficult to envisage the form of the nonlinearity and little can be anticipated about the optimal compression scheme. The theory allows us, however, to find the weighting function and the optimal nonlinearity by means of numerical integration techniques. The shape of the nonlinearity is depicted in the bottom-left panel of figure 3. We see that the optimal nonlinearity computed in the monoterminal case is different from that obtained in a decentralized system with $N = 2$ terminals, and also different from the optimal nonlinearity computed by exploiting the technique based on Van Trees bound. However, these differences are usually moderate. The rate distortion curves corresponding to $N = 2$ are shown in figure 6.

D. Correlated Gaussian Case

Previous examples do confirm that appropriate quantizers can be specifically designed for the estimation task. However, they may also suggest that such a design could be more or less equivalently performed as if a monoterminal ($N = 1$) problem were in force. As a matter of fact, in the cases addressed there are either no differences between $N = 1, 2, \ldots, \infty$ (see Sects. V-A and V-B) or these differences are moderate, as in the case treated in Sect. V-C (see the bottom-left panel of figure 3).

\footnote{We can explicitly compute the optimal MMSE estimator in the case of $N = 2$ sensors – the resulting analytical expression is unpleasant, so we do not report it here.}

We now consider a case in which a marked breakpoint exists between the mono- and multi-terminal solutions. Consider hence the decentralized problem of estimating the correlation coefficient $\theta$ between two Gaussian random variables$^{10}$, where the joint distribution of the observations, conditioned on $\theta$, is:

$$f(x_1, x_2|\theta) = \frac{1}{2\pi\sqrt{1-\theta^2}} \exp \left\{ -\frac{x_1^2 + x_2^2 - 2\theta x_1 x_2}{2(1-\theta^2)} \right\}.$$  

We assume also that the priori pdf on $\theta$ is $f_\theta(\theta) \propto (1 - \theta^2)^{-3/2}, \theta \in [-\theta^*, \theta^*]$ and zero otherwise, with $0 < \theta^* < 1$. Such a probability density, which turns out to be of a certain interest in some Bayesian frameworks [16], appropriately enhances large (positive/negative) correlations between sensors.

The MMSE estimator, as well as the function and quantities relevant for the evaluation of $g(\cdot)$, are not easy to compute, and we thus resort to numerical integration; the optimization based on Van Trees is inappropriate, due to the lack of conditional independence between sensors. The nonlinearity found is displayed in bottom-right panel of figure 3. Figure 7 depicts the rate distortion region for the estimation of the correlation coefficient, with $N = 2$.

VI. Summary

We have investigated quantized multi-terminal estimation of a common random variable under the assumptions that

- the quantizers are noncooperative,
- quantization is high-rate,
- quantization at each sensor is scalar and uses a companding model, and
- data is transmitted from sensors to fusion center via optimal Slepian-Wolf coding.

\footnote{In this case the mono-terminal posing makes no sense at all.}
We build on Poor’s work on optimal high-rate quantization for detection and estimation. However, the last bullet (joint optimal noiseless source coding) is quite different from [20], and exerts a marked effect.

We have derived an expression for the rate-distortion function, where distortion is MSE in terms of the inference task. An expression for the rate-distortion optimal performance is easy, and we have an explicit expression for the optimal companding nonlinearity. For N-sensors this latter involves (N – 1)-dimensional integrals, and hence it may not be practical except for small N. Accordingly we develop an approximation based on substitution of the Fisher information for MSE, and we find a simple way to optimize this proxy.

Estimation of a common Gaussian-distributed mean among sensors with additive Gaussian noise is known as the CEO problem. For the CEO case our scheme optimally employs a trivial (linear) companding function, and we find that the loss from use of such scalar quantization is 1.53 dB versus the bound of Zamir and Berger. We also look at richer inference problems than the case, and find more interesting compander forms as well as significant variability in the optimal companding nonlinearity as a function of N.

APPENDIX

The Van Trees bound [25] provides a lower limit for the MSE for estimation of a random variable θ from a collection of data x,

\[ \text{MSE} \geq \frac{1}{E \left[ \left( \frac{\partial \ln f_\theta(\theta)}{\partial \theta} \right)^2 \right] + E \left[ \left( \frac{\partial \ln f(x|\theta)}{\partial \theta} \right)^2 \right]}. \tag{41} \]

If the exact evaluation of eq. (26) is challenging, this bound could be profitably exploited as an alternative distortion measure. Accordingly, we basically repeat the derivation of section III-B using this new distortion proxy, and also assuming observations conditionally independent given θ.

Let us define the scores of unquantized and quantized observations, respectively:

\[ \xi_x(x, \theta) = \frac{\partial \ln f(x|\theta)}{\partial \theta}, \quad \xi_q(q, \theta) = \frac{\partial \ln p(q|\theta)}{\partial \theta}, \]

and the associated statistical average with respect to both x and θ

\[ I_x = E \left[ \xi_x^2(x, \theta) \right], \quad I_q = E \left[ \xi_q^2(q, \theta) \right]. \]

Further, let \( I_\theta \) be the first addend in the numerator of the RHS of eq. (41), which is the term accounting for the priori information about θ. The estimation errors for unquantized and quantized data may be approximated, respectively, as:

\[ D_{\text{min}} \approx \frac{1}{I_\theta + N I_x}, \quad D \approx \frac{1}{I_\theta + N I_q}. \tag{42} \]

It is expedient to rewrite the score from quantized samples as

\[ \xi_q(q, \theta) = \frac{\partial p(q|\theta)/\partial \theta}{p(q|\theta)} = \frac{\int_{x \in \mathbb{R}_n} \partial f(x|\theta)/\partial \theta \, dx}{\int_{x \in \mathbb{R}_n} f(x|\theta) \, dx}, \]

\[ = \int_{x \in \mathbb{R}_n} \xi_x(x, \theta)f(x|\theta) \, dx, \]

which reveals that \( \xi_q(q, \theta) \) is the centroid of the unquantized score \( \xi_x(x, \theta) \), computed under the conditional pdf \( f(x|\theta) \). Then, we have (score arguments omitted for simplicity)

\[ E_{x|\theta} \left[ \xi_x^2 \right] = E_x|\theta \left[ (\xi_x - \xi_q)^2 \right] + E_{x|\theta} \left[ \xi_q^2 \right], \tag{43} \]

where we have applied the orthogonality principle (\( \xi_x - \xi_q \) ⊥ \( \xi_q \)), the centroid being an MMSE estimate of \( \xi_x \) given the partition region [8], for any prescribed value of \( \theta \). Further averaging over \( \theta \) yields

\[ I_q = I_x - E \left[ (\xi_x - \xi_q)^2 \right]. \tag{44} \]

Putting this latter in the rightmost form of eq. (42), and using \((1 - \epsilon)^{-1} c^{-\epsilon} \approx (1 + \epsilon)\), allows rewriting the distortion as:

\[ D \approx \frac{1}{I_\theta + N I_x} + \frac{N}{(I_\theta + N I_q)^2} E \left[ (\xi_x - \xi_q)^2 \right]. \tag{45} \]

Now, eq. (45) is reminiscent of eq. (11), and as before we have to minimize a reproduction error. To this aim, note that \( \xi_q \) is the centroid of \( \xi_x \), so that, paralleling the development of section III-B, we can approximate \( \xi_x(x, \theta) - \xi_x(\nu, \theta) \approx (\partial \xi_x(x, \theta)/\partial x)(x - \nu) \) and, by the same arguments as those following eq. (13), again exchange the points \( \nu \) with the midpoints of the partition regions. This implies

\[ E \left[ (\xi_x - \xi_q)^2 \right] = \int \int \left( \frac{\partial \xi_x(x, \theta)}{\partial x} \right)^2 (x - \nu)^2 f(x, \theta) \, dx \, d\theta. \tag{46} \]

If we now set

\[ w^2(x) = \frac{1}{(I_\theta + N I_x)^2} \int \left( \frac{\partial^2 \ln f(x|\theta)}{\partial x \partial \theta} \right)^2 f_\theta(\theta|x) \, d\theta, \tag{47} \]
eq. (46) takes the form

\[ E \left[ (\xi_x - \xi_q)^2 \right] = (I_0 + N I_x)^2 \int w^2(x)(x-\nu)^2 f(x)dx. \]  

\[ (48) \]

Further substituting in eq. (45), and accounting for the the expression of \( D_{\text{min}} \) given in eq. (42), yields

\[ D \approx D_{\text{min}} + N \int w^2(x)(x-\nu)^2 f(x)dx, \]

\[ (49) \]

which is easily recognized to reproduce eq. (16). Consequently, along the same lines of the derivation in sect. III-B, we finally obtain

\[ D = D_{\text{min}} + N(\delta^2/12) \int f(s)|w(s)\gamma(s)|^2 \int dsd\theta, \]

\[ (50) \]

which is nothing but eq. (25). Accordingly, the optimization yields results identical to those of sect. III-C, all being unchanged in terms of rate, see eq. (24).

REFERENCES


