MMSE Dimension

Yihong Wu
Department of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: yihongwu@princeton.edu

Sergio Verdú
Department of Electrical Engineering
Princeton University
Princeton, NJ 08544, USA
Email: verdu@princeton.edu

***Abstract*—If N is standard Gaussian, the minimum mean-square error (MMSE) of estimating X based on $\sqrt{\mathsf{snr}}X+N$ vanishes at least as fast as $\frac{1}{\mathsf{snr}}$ as $\mathsf{snr}\to\infty$. We define the *MMSE dimension* of X as the limit as $\mathsf{snr}\to\infty$ of the product of snr and the MMSE. For discrete, absolutely continuous or mixed X we show that the MMSE dimension equals Rényi's information dimension. However, for singular X, we show that the product of snr and MMSE oscillates around information dimension periodically in snr (dB). We also show that discrete side information does not reduce MMSE dimension. These results extend considerably beyond Gaussian N under various technical conditions.**

I. Introduction

A. Basic Setup

The minimum mean square error (MMSE) plays a pivotal role in estimation theory and Bayesian statistics. Due to the lack of closed-form expressions for posterior distributions and conditional expectations, exact MMSE formulae are scarce. Asymptotic analysis is more tractable and sheds important insights about how the fundamental estimation-theoretic limits depend on the input and noise statistics. The theme of this paper is the high-SNR scaling law of the MMSE of estimating X based on $\sqrt{\mathsf{snr}}X+N$ when N is independent of X.

The MMSE of estimating X based on Y is given by

$$\mathsf{mmse}(X|Y)=\mathbb{E}[(X-\mathbb{E}[X|Y])^2]. \tag{1}$$

When Y is related to X through an additive-noise channel with gain $\sqrt{\mathsf{snr}}$, i.e.,

$$Y=\sqrt{\mathsf{snr}}X+N \tag{2}$$

where $N\perp\!\!\!\perp X$, we denote

$$\mathsf{mmse}(X,N,\mathsf{snr})=\mathsf{mmse}(X|\sqrt{\mathsf{snr}}X+N), \tag{3}$$

and, in particular, when the noise is Gaussian (denoted by N_{G}), we simplify

$$\mathsf{mmse}(X,\mathsf{snr})=\mathsf{mmse}(X,N_{\mathsf{G}},\mathsf{snr}). \tag{4}$$

B. Asymptotics of MMSE

The low-SNR asymptotics of $\mathsf{mmse}(X,\mathsf{snr})$ has been studied extensively in [1, Section II.F], where the Taylor expansion of $\mathsf{mmse}(X,\mathsf{snr})$ at $\mathsf{snr}=0$ has been obtained and the coefficients turn out to depend only on the moments of X. However, the asymptotics in the high-SNR regime have not received much attention in the literature. The high-SNR behavior depends on the input distribution: For example, for binary X, $\mathsf{mmse}(X,\mathsf{snr})$ decays exponentially, while for standard Gaussian X,

$$\mathsf{mmse}(X,\mathsf{snr})=\frac{1}{\mathsf{snr}+1}. \tag{5}$$

Unlike the low-SNR regime, the high-SNR asymptotics depend on the measure-theoretical structure of the input distribution rather than its moments.

Before defining MMSE dimension, note that

$$0\le\mathsf{mmse}(X,N,\mathsf{snr})\le\frac{\mathsf{var}N}{\mathsf{snr}}, \tag{6}$$

where the upper bound can be achieved using the affine estimator $f(y)=\frac{y-\mathbb{E}N}{\sqrt{\mathsf{snr}}}$. Assuming N has finite variance, as $\mathsf{snr}\to\infty$, we have[1]:

$$\mathsf{mmse}(X,N,\mathsf{snr})=\mathrm{O}\left(\frac{1}{\mathsf{snr}}\right) \tag{7}$$

Seeking a finer characterization, we are interested in the exact scaling constant in (7). To this end, we define the *upper* and *lower MMSE dimension* of the pair (X,N) as:

$$\overline{\mathscr{D}}(X,N)=\limsup_{\mathsf{snr}\to\infty}\frac{\mathsf{snr}\cdot\mathsf{mmse}(X,N,\mathsf{snr})}{\mathsf{var}N}, \tag{8}$$

$$\underline{\mathscr{D}}(X,N)=\liminf_{\mathsf{snr}\to\infty}\frac{\mathsf{snr}\cdot\mathsf{mmse}(X,N,\mathsf{snr})}{\mathsf{var}N}. \tag{9}$$

If $\overline{\mathscr{D}}(X,N)=\underline{\mathscr{D}}(X,N)$, the common value is denoted by $\mathscr{D}(X,N)$, called the *MMSE dimension* of (X,N). This information measure governs the high-SNR scaling law of MMSE and sharpens (7) to:

$$\mathsf{mmse}(X,N,\mathsf{snr})=\frac{\mathscr{D}(X,N)\mathsf{var}N}{\mathsf{snr}}+\mathrm{o}\left(\frac{1}{\mathsf{snr}}\right). \tag{10}$$

When the side information U (independent of N) is present, replacing $\mathsf{mmse}(X,N,\mathsf{snr})$ by $\mathsf{mmse}(X,N,\mathsf{snr}|U)$, the *conditional MMSE dimension* of (X,N) given U is similarly defined, denoted by $\mathscr{D}(X,N|U)$. When N is Gaussian, we use the notation $\mathscr{D}(X)$ ($\mathscr{D}(X|U)$), called the *(conditional) MMSE dimension* of X. By (6), we have

$$0\le\underline{\mathscr{D}}(X,N|U)\le\overline{\mathscr{D}}(X,N|U)\le 1. \tag{11}$$

[1]We use the following asymptotic notation: $f(x)=\mathrm{O}\left(g(x)\right)$ if and only if $\limsup\frac{|f(x)|}{|g(x)|}<\infty$. $f(x)=\Omega(g(x))$ if and only if $g(x)=\mathrm{O}\left(f(x)\right)$. $f(x)=\Theta(g(x))$ if and only if $f(x)=\mathrm{O}\left(g(x)\right)$ and $f(x)=\Omega(g(x))$

Although most of our attention is focused on square integrable noise, $\mathsf{mmse}(X,N,\mathsf{snr})$ is finite even for infinite-variance N, as long as $\mathsf{var}X<\infty$. But (8) and (9) no longer make sense in that case. Hence the scaling law in (7) could fail. Examples can be constructed where $\mathsf{mmse}(X,N,\mathsf{snr})$ decays strictly slower than $\frac{1}{\mathsf{snr}}$, *e.g.*, as $\frac{\log \mathsf{snr}}{\mathsf{snr}}$ or even $\frac{1}{\sqrt{\mathsf{snr}}}$. However, for certain X, $\mathsf{mmse}(X,N,\mathsf{snr})$ can still decay according to $\frac{1}{\mathsf{snr}}$ for some infinite-variance N [2].

C. Asymptotics of Fisher information

In the special case of Gaussian noise it is interesting to draw conclusions on the asymptotic behavior of Fisher's information based on our results. Similarly to the MMSE dimension, we can define the *Fisher dimension* of a random variable X as follows:

$$\overline{\mathscr{J}}(X)=\limsup_{\epsilon\downarrow 0}\epsilon^2\cdot J(X+\epsilon N_{\mathsf{G}}), \tag{12}$$

$$\underline{\mathscr{J}}(X)=\liminf_{\epsilon\downarrow 0}\epsilon^2\cdot J(X+\epsilon N_{\mathsf{G}}). \tag{13}$$

In view of the representation of MMSE via the Fisher information of the channel output [3, (1.3.4)], [1, (58)]:

$$\mathsf{snr}\cdot\mathsf{mmse}(X,\mathsf{snr})=1-J(\sqrt{\mathsf{snr}}X+N_{\mathsf{G}}) \tag{14}$$

and $J(aZ)=a^{-2}J(Z)$, we conclude that Fisher dimension and MMSE dimension are complementary of each other:

$$\overline{\mathscr{J}}(X)+\underline{\mathscr{D}}(X)=\underline{\mathscr{J}}(X)+\overline{\mathscr{D}}(X)=1. \tag{15}$$

D. Connections to asymptotic statistics

The high-SNR asymptotic behavior of $\mathsf{mmse}(X,\mathsf{snr})$ is equivalent to the behavior of the Bayesian risk for the Gaussian location model in the large sample limit, where P_X is the prior distribution and the sample size n plays the role of snr. Let $\{N_i:i\in\mathbb{N}\}$ be a sequence of i.i.d. standard Gaussian random variables independent of X and denote $Y_i=X+N_i$. By the sufficiency of the sample mean $\bar{Y}=\frac{1}{n}\sum_{i=1}^n Y_i$ in Gaussian location models, we have

$$\mathsf{mmse}(X|Y^n)=\mathsf{mmse}(X|\bar{Y})=\mathsf{mmse}(X,n). \tag{16}$$

Therefore as sample size grows, the Bayesian risk of estimating X vanishes as $\mathrm{O}\left(\frac{1}{n}\right)$ with the scaling constant given by the MMSE dimension of the prior:

$$\mathscr{D}(X)=\lim_{n\to\infty}\mathsf{mmse}(X|Y^n)\,n. \tag{17}$$

The asymptotics of $\mathsf{mmse}(X|Y^n)$ has been studied in [4], [5] for absolutely continuous priors and general models where X and Y are not necessarily related by additive Gaussian noise. Those results are compared with ours in Section II.

II. Main Results

A. Connections to information dimension

The MMSE dimension is intimately related to the *information dimension* defined by Rényi in [6]:

Definition 1. Let X be a real-valued random variable. Denote for a positive integer m the quantized version of X:

$$\langle X\rangle_m=\frac{\lfloor mX\rfloor}{m}. \tag{18}$$

Define[2]

$$\underline{d}(X)=\liminf_{m\to\infty}\frac{H\left(\langle X\rangle_m\right)}{\log m} \tag{19}$$

and

$$\overline{d}(X)=\limsup_{m\to\infty}\frac{H\left(\langle X\rangle_m\right)}{\log m}, \tag{20}$$

where $\underline{d}(X)$ and $\overline{d}(X)$ are called *lower* and *upper information dimensions* of X respectively. If $\underline{d}(X)=\overline{d}(X)$, the common value is called the *information dimension* of X, denoted by $d(X)$.

The following theorem reveals a connection between the MMSE dimension and the information dimension of the input:

Theorem 1. *If $\mathbb{E}[X^2]<\infty$, then*

$$\underline{\mathscr{D}}(X)\le\underline{d}(X)\le\overline{d}(X)\le\overline{\mathscr{D}}(X), \tag{21}$$

Therefore, if $\mathscr{D}(X)$ exists, then $d(X)$ exists, and

$$\mathscr{D}(X)=d(X). \tag{22}$$

Due to space limitations, proofs are referred to [2].

In [7, p.755] it is conjectured that $\underline{\mathscr{J}}(X)=1-\overline{d}(X)$, or equivalently, in view of (15), $\overline{\mathscr{D}}(X)=\overline{d}(X)$. This holds for distributions without singular components but not in general. The Cantor distribution gives a counterexample to the general conjecture (see Section II-D). However, whenever X has a discrete-continuous mixed distribution, (22) holds and information dimension governs the high-SNR asymptotics of MMSE.

Next we drop the assumption of $\mathsf{var}X<\infty$ and proceed to give results for various input and noise distributions.

B. Absolutely continuous inputs

Theorem 2. *If X has an absolutely continuous distribution with respect to Lebesgue measure, then $\mathscr{D}(X)=1$, i.e.,*

$$\mathsf{mmse}(X,\mathsf{snr})=\frac{1}{\mathsf{snr}}+\mathrm{o}\left(\frac{1}{\mathsf{snr}}\right). \tag{23}$$

In view of (40), Theorem 2 implies that absolutely continuous priors result in procedures whose asymptotic Bayes risk coincides with the minimax level.

Under certain regularity conditions on N (X resp.) we can extend Theorem 2 to show

$$\mathscr{D}(X,N)=1 \tag{24}$$

[2]Throughout the paper, natural logarithms are adopted and information units are nats.

for all absolutely continuous X (N resp.). For example, when X has a continuous and bounded density, (24) holds for all square-integrable N even without a density.

A refinement of Theorem 2 entails the second-order expansion for $\mathsf{mmse}(X,\mathsf{snr})$ for absolutely continuous input X. This involves the Fisher information of X. Suppose $J(X)<\infty$, then $J(\sqrt{\mathsf{snr}}X+N_{\mathsf{G}})\le J(X)\mathsf{snr}^{-1}$, by the convexity and translation invariance of Fisher information. In view of (14), we have

$$\mathsf{mmse}(X,\mathsf{snr})=\frac{1}{\mathsf{snr}}+\mathrm{O}\left(\frac{1}{\mathsf{snr}^2}\right), \tag{25}$$

which is a slight improvement of (23).

Under stronger conditions the second-order term can be determined exactly. A result in [8] states that: if $J(X)<\infty$ and the density of X satisfies certain regularity conditions [8, (3) – (7)], then

$$J(X+\epsilon N_{\mathsf{G}})=J(X)+\mathrm{O}(\epsilon). \tag{26}$$

Therefore in view of (14), we have

$$\mathsf{mmse}(X,\mathsf{snr})=\frac{1}{\mathsf{snr}}-\frac{J(X)}{\mathsf{snr}^2}+\mathrm{o}\left(\frac{1}{\mathsf{snr}^2}\right). \tag{27}$$

This result can be understood as follows: using (14) and Stam's inequality [9], we have [10, pp. 72–73]

$$\mathsf{mmse}(X,\mathsf{snr})\ge\frac{1}{J(X)+\mathsf{snr}}, \tag{28}$$

which is also known as the Bayesian Cramér-Rao bound (or the Van Trees inequality). In view of (27), we see that (28) is asymptotically tight for sufficiently regular densities of X.

Instead of using the asymptotic expansion of Fisher information and Stam's inequality, we can show that (27) holds for a much broader class of densities of X and non-Gaussian noise: if X has finite third moment and its density has bounded first two derivatives, then for any finite-variance N,

$$\mathsf{mmse}(X,N,\mathsf{snr})=\frac{\mathsf{var}N}{\mathsf{snr}}-\frac{J(X)\mathsf{var}^2N}{\mathsf{snr}^2}+\mathrm{o}\left(\frac{1}{\mathsf{snr}^2}\right). \tag{29}$$

This equation reveals a new *operational role* of $J(X)$. The regularity conditions imposed on the input density are much weaker and easier to check than those in [8]; however, (27) is slightly stronger than (29) because the $\mathrm{o}\left(\mathsf{snr}^{-2}\right)$ term in (27) is in fact $\mathrm{O}\left(\mathsf{snr}^{-5/2}\right)$, as a result of (26).

It is interesting to compare (29) to the asymptotic expansion of Bayesian risk in the large sample limit [5]. Under the regularity conditions in [5, Theorem 5.1a], we have

$$\mathsf{mmse}(X|Y_1^n)=\frac{1}{n\,J(N)}-\frac{J(X)}{n^2\,J(N)^2}+\mathrm{o}\left(\frac{1}{n^2}\right). \tag{30}$$

On the other hand, by (29) we have

$$\mathsf{mmse}(X|\bar{Y})=\mathsf{mmse}(X,n)=\frac{\mathsf{var}N}{n}-\frac{J(X)(\mathsf{var}N)^2}{n^2}+\mathrm{o}\left(\frac{1}{n^2}\right) \tag{31}$$

When N is Gaussian, (31) agrees with (30), but our proof requires much weaker regularity conditions. When N is not Gaussian, (31) is inferior to (30) on the first term, due to the Cramér-Rao bound $\mathsf{var}N\ge\frac{1}{J(N)}$. This agrees with the fact that sample mean is asymptotically suboptimal for non-Gaussian noise, and the suboptimality is characterized by the gap in the Cramér-Rao inequality.

To conclude the discussion of absolutely continuous inputs, we give an example where (29) fails:

Example 1. Consider X and N uniformly distributed on $[0,1]$. It can be shown that $\mathsf{mmse}(X,N,\mathsf{snr})=\mathsf{var}N\left(\frac{1}{\mathsf{snr}}-\frac{1}{2\mathsf{snr}^{\frac{3}{2}}}\right)$ for $\mathsf{snr}\ge 4$. Note that (29) does not hold because X does not have a differentiable density, hence $J(X)=\infty$ (in the sense of the generalized Fisher information in [11, Definition 4.1]).

C. Conditional MMSE dimension

Next we present results for general mixed distributions, which can be equivalently stated via the conditional MMSE dimension.

Theorem 3 (Conditional MMSE dimension)**.**

$$\overline{\mathscr{D}}(X,N)\ge\overline{\mathscr{D}}(X,N|U), \tag{32}$$

$$\underline{\mathscr{D}}(X,N)\ge\underline{\mathscr{D}}(X,N|U). \tag{33}$$

When N is Gaussian and U is discrete, (32) and (33) hold with equality.

Theorem 3 implies that any *discrete* side information does not change the high-SNR asymptotics of MMSE. Consequently, knowing arbitrarily finitely many digits of X does not reduce its MMSE dimension. Another application of Theorem 3 is to determine the MMSE dimension of inputs with mixed distributions, which are frequently used in statistical models of sparse signals [12], [13], [14]. According to Theorem 3, knowing the support does not decrease the MMSE dimension.

Corollary 1. *Let $X=UZ$ where U is independent of Z, taking values in $\{0,1\}$ with $\mathbb{P}\{U=1\}=\rho$. Then $\overline{\mathscr{D}}(X)=\rho\overline{\mathscr{D}}(Z)$, $\underline{\mathscr{D}}(X)=\rho\underline{\mathscr{D}}(Z)$.*

Having proved results about countable mixtures, the following theorem about MMSE dimension of discrete random variables is just a simple corollary.

Theorem 4. *For X discrete (even with countably infinite alphabet), $\mathscr{D}(X)=0$.*

Without recourse to conditional MMSE dimension, Theorem 4 can be extended to $\mathscr{D}(X,N)=0$ for all discrete X and absolutely continuous N.

Obtained by combining Theorems 2 – 4, the next result gives a complete characterization to the MMSE dimension of discrete-continuous mixtures. Together with Theorem 1, Theorem 5 also provides an MMSE-based proof of Rényi's theorem on information dimension [6] for square integrable inputs.

Theorem 5. *If X is distributed according to a mixture of an absolutely continuous and a discrete distribution, then its MMSE dimension equals the weight of the absolutely continuous part.*

In [2] we generalize Theorem 5 to non-Gaussian noise.

For square integrable input, the MSE attained by the optimal linear estimator is $\frac{\mathsf{var}N}{\mathsf{snr}}+\mathrm{o}\left(\frac{1}{\mathsf{snr}}\right)$. Therefore the linear estimator is dimensionally optimal for estimating absolutely continuous random variables contaminated by additive Gaussian noise, in the sense that it achieves the input MMSE dimension.

To conclude this subsection, we illustrate Theorem 5 by the following examples:

Example 2 (Continuous input)**.** If $X \sim \mathcal{N}(0,1)$, then $\mathsf{mmse}(X,\mathsf{snr})$ is given in (5) and Theorem 2 holds.

Example 3 (Discrete input)**.** If X is equiprobable on $\{-1,1\}$, then

$$\mathsf{mmse}(X,\mathsf{snr}) = \mathrm{O}\left(\frac{1}{\sqrt{\mathsf{snr}}}\mathrm{e}^{-\frac{\mathsf{snr}}{2}}\right) \tag{34}$$

and therefore $\mathscr{D}(X) = 0$ as predicted by Theorem 4. In fact, exponential decay holds for all inputs whose alphabet has no accumulation points. Otherwise, the MMSE can decay polynomially [2].

Example 4 (Mixed input)**.** Let N be uniformly distributed in $[0,1]$, and let X be distributed according to an equal mixture of a mass at 0 and a uniform distribution on $[0,1]$. Then $\mathsf{mmse}(X,N,\mathsf{snr}) = \mathsf{var}N\left(\frac{1}{2\mathsf{snr}}+\frac{1}{4\mathsf{snr}^{\frac{3}{2}}}\right)+\mathrm{o}\left(\frac{1}{\mathsf{snr}^{\frac{3}{2}}}\right)$, which implies $\mathscr{D}(X,N) = \frac{1}{2}$ as predicted by Theorem 5 for non-Gaussian noise.

D. Singularly continuous inputs

We focus on a special class of inputs with self-similar singular distributions [15, p. 36]: inputs with i.i.d. digits.

Theorem 6. *Let $X \in [0,1]$ a.s., whose M-ary expansion $X = \sum_{j\in\mathbb{N}}(X)_j M^{-j}$ consists of i.i.d. digits $\{(X)_j\}$ with common distribution P. Then for any finite-variance N, there exists a $2\log M$-periodic function $\Phi_{X,N}:\mathbb{R}\to[0,1]$, such that as $\mathsf{snr}\to\infty$,*

$$\mathsf{mmse}(X,N,\mathsf{snr}) = \frac{\mathsf{var}N}{\mathsf{snr}}\Phi_{X,N}(\log\mathsf{snr})+\mathrm{o}\left(\frac{1}{\mathsf{snr}}\right). \tag{35}$$

The lower and upper MMSE dimension of (X,N) are given by

$$\underline{\mathscr{D}}(X,N) = \limsup_{b\to\infty}\Phi_{X,N}(b) = \sup_{0\le b\le 2\log M}\Phi_{X,N}(b), \tag{36}$$

$$\overline{\mathscr{D}}(X,N) = \liminf_{b\to\infty}\Phi_{X,N}(b) = \inf_{0\le b\le 2\log M}\Phi_{X,N}(b). \tag{37}$$

Moreover, when $N = N_{\mathsf{G}}$ is Gaussian, the average of $\Phi_{X,N_{\mathsf{G}}}$ over one period coincides with the information dimension of X:

$$\frac{1}{2\log M}\int_0^{2\log M}\Phi_{X,N_{\mathsf{G}}}(b)\mathrm{d}b = d(X) = \frac{H(P)}{\log M}. \tag{38}$$

Theorem 6 shows that in the high-SNR regime, the function $\mathsf{snr}\,\mathsf{mmse}(X,N,\mathsf{snr})$ is periodic in snr (dB) with period M^2 (dB). This periodicity arises from the *self-similarity* of the input, and unlike the lower and upper dimension, the period depends on the input distribution but not the noise. Although Theorem 6 reveals the oscillatory nature of $\mathsf{mmse}(X,N,\mathsf{snr})$, we do not have a general formula to compute the lower (or upper) MMSE dimension of (X,N). However, when the noise is Gaussian, Theorem 1 provides a sandwich bound in terms of the information dimension of X, which is reconfirmed by combining (36) – (38).

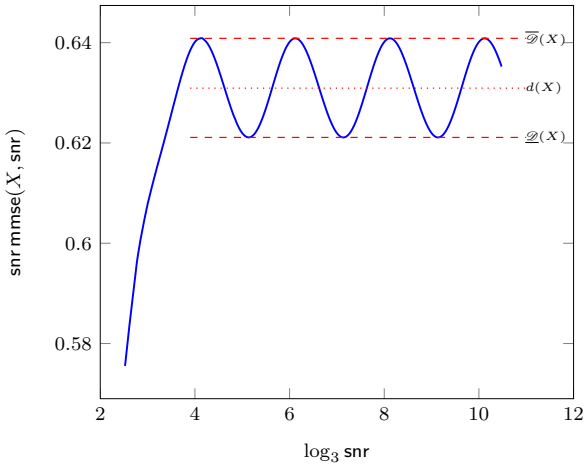

Fig. 1. $\mathsf{snr}\,\mathsf{mmse}(X,\mathsf{snr})$ when X has a ternary Cantor distribution.

As an illustrative example, consider Cantor-distributed X whose ternary expansion consists of i.i.d. digits, and for each j, $\mathbb{P}\{(X)_j=0\} = \mathbb{P}\{(X)_j=2\} = 1/2$. According to Theorem 6, in the high-SNR regime, $\mathsf{snr}\,\mathsf{mmse}(X,\mathsf{snr})$ oscillates periodically in $\log\mathsf{snr}$ with period $2\log 3$, as plotted in Fig. 1. The lower and upper MMSE dimensions of the Cantor distribution turn out to be (to six decimals): $\underline{\mathscr{D}}(X) = 0.621102$, $\overline{\mathscr{D}}(X) = 0.640861$. The information dimension $d(X) = \log_3 2 = 0.630930$ is sandwiched between $\underline{\mathscr{D}}(X)$ and $\overline{\mathscr{D}}(X)$, according to Theorem 1. From this and other numerical evidence it is tempting to conjecture that when the noise is Gaussian,

$$d(X) = \frac{\underline{\mathscr{D}}(X)+\overline{\mathscr{D}}(X)}{2}. \tag{39}$$

It should be pointed out that the sandwich bounds in (21) need not hold when N is not Gaussian. For example, for X Cantor distributed and N uniformly distributed in $[0,1]$, numerical calculation shows that $d(X) = \log_3 2 > \overline{\mathscr{D}}(X,N) = 0.5741$.

III. Concluding Remarks

Through the high-SNR asymptotics of MMSE in Gaussian channels, we defined a new information measure called MMSE dimension. Although stemming from estimation-theoretic principles, MMSE dimension shares several important features with Rényi's information dimension. By Theorem 3, MMSE dimensions are affine functionals, a fundamental property shared by information dimension [12, Theorem 2]. According to Theorem 1, information dimension is sandwiched between the lower and upper MMSE dimensions. For distributions with no singular components, they coincide to be the weight on the

absolutely continuous part of the distribution. In [12], we have shown that the information dimension plays a pivotal role in almost lossless analog compression, an information theory for noiseless compressed sensing. In fact we have shown that the MMSE dimension serves as the fundamental limit in noisy compressed sensing with stable recovery.

Via the input-noise duality

$$\mathsf{snr} \cdot \mathsf{mmse}(X, N, \mathsf{snr}) = \mathsf{mmse}\left(N, X, \mathsf{snr}^{-1}\right) \tag{40}$$

and the following equivalent definition of the MMSE dimension

$$\mathscr{D}(X, N) = \lim_{\epsilon \downarrow 0} \mathsf{mmse}(N, X, \epsilon), \tag{41}$$

studying high-SNR asymptotics provides insights into the behavior of $\mathsf{mmse}(X, N, \mathsf{snr})$ for non-Gaussian noise N. Combining various results from Sections II, we observe that unlike in Gaussian channels where $\mathsf{mmse}(X, \mathsf{snr})$ is decreasing and convex in snr, $\mathsf{mmse}(X, N, \mathsf{snr})$ can behave very irregularly in general. To illustrate this, we consider the case where standard Gaussian input is contaminated by various additive noises. For all N, it is evident that $\mathsf{mmse}(X, N, 0) = \mathsf{var}X = 1$. The behavior of MMSE associated with Gaussian, Bernoulli and Cantor distributed noises is (Fig. 2)

- For standard Gaussian N, $\mathsf{mmse}(X, \mathsf{snr}) = \frac{1}{1+\mathsf{snr}}$ is *continuous* at $\mathsf{snr} = 0$ and decreases monotonically according to $\frac{1}{\mathsf{snr}}$. This monotonicity is due to the MMSE data-processing inequality and the stability of Gaussian distribution.
- For equiprobable Bernoulli N, $\mathsf{mmse}(X, N, \mathsf{snr})$ is *discontinuous* at $\mathsf{snr} = 0$. As $\mathsf{snr} \to 0$, the MMSE vanishes according to $\mathrm{O}\left(\frac{1}{\sqrt{\mathsf{snr}}} \mathrm{e}^{-\frac{1}{2\mathsf{snr}}}\right)$, in view of (40) and (34), and since it also vanishes as $\mathsf{snr} \to \infty$, it is not monotonic with $\mathsf{snr} > 0$.
- For Cantor distributed[3] N, $\mathsf{mmse}(X, N, \mathsf{snr})$ is also *discontinuous* at $\mathsf{snr} = 0$. According to Theorem 6, as $\mathsf{snr} \to 0$, the MMSE oscillates relentlessly between $\overline{\mathscr{D}}(N)$ and $\underline{\mathscr{D}}(N)$ and does not have a limit (See the zoom-in plot in Fig. 3).

Nevertheless, as $\mathsf{snr} \to \infty$, the MMSE vanishes as $\frac{\mathsf{var}N}{\mathsf{snr}}$ regardless of the noise.

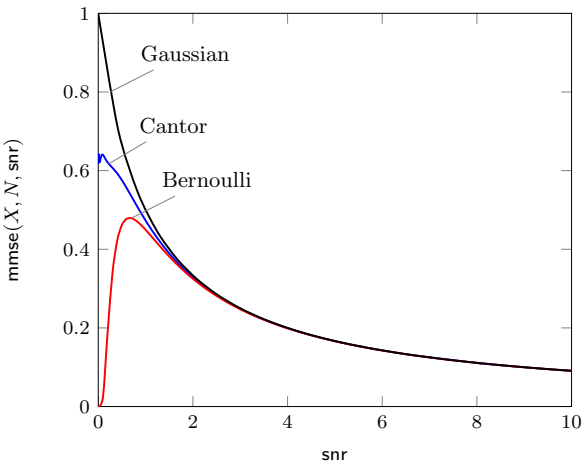

Fig. 2. $\mathsf{mmse}(X, N, \mathsf{snr})$ when X is standard Gaussian and N is standard Gaussian, equiprobable Bernoulli or Cantor distributed.

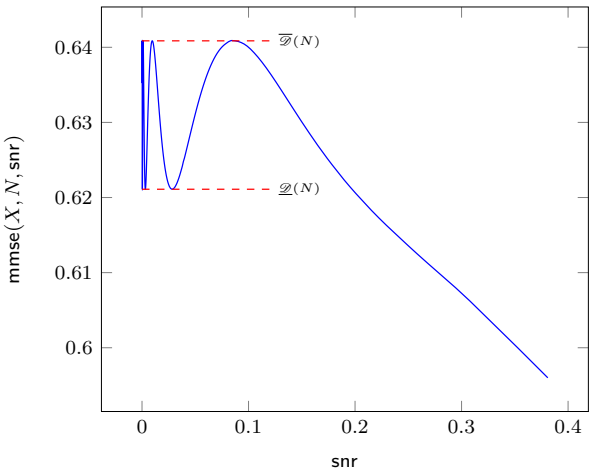

Fig. 3. $\mathsf{mmse}(X, N, \mathsf{snr})$ when X is standard Gaussian and N is Cantor distributed.

REFERENCES

[1] D. Guo, S. Shamai, and S. Verdú, "Mutual Information and Minimum Mean-Square Error in Gaussian Channels," *IEEE Transactions on Information Theory*, vol. 51, no. 4, pp. 1261 – 1283, April 2005.

[2] Y. Wu and S. Verdú, "MMSE Dimension," submitted to *IEEE Transactions on Information Theory*, 2009.

[3] L. D. Brown, "Admissible estimators, recurrent diffusions, and insoluble boundary value problems," *The Annals of Mathematical Statistics*, vol. 42, no. 3, pp. 855–903, 1971.

[4] M. V. Burnašev, "Investigation of second order properties of statistical estimators in a scheme of independent observations," *Math. USSR Izvestiya*, vol. 18, no. 3, pp. 439–467, 1982.

[5] J. K. Ghosh, *Higher Order Asymptotics*. Institute of Mathematical Statistics, 1994.

[6] A. Rényi, "On the dimension and entropy of probability distributions," *Acta Mathematica Hungarica*, vol. 10, no. 1 – 2, March 1959.

[3]Since the standard Cantor distribution has variance $\frac{1}{8}$, N has been scaled by $2\sqrt{2}$ to make it unit-variance.

[7] A. Guionnet and D. Shlyakhtenko, "On classical analogues of free entropy dimension," *Journal of Functional Analysis*, vol. 251, no. 2, pp. 738 – 771, 2007.

[8] V. V. Prelov and E. C. van der Meulen, "Asymptotics of Fisher Information under Weak Perturbation," *Problems of Information Transmission*, vol. 31, no. 1, pp. 14 – 22, 1995.

[9] A. Stam, "Some inequalities satisfied by the quantities of information of Fisher and Shannon," *Information and Control*, vol. 2, no. 2, pp. 101–112, 1959.

[10] H. Van Trees, *Detection, Estimation, and Modulation theory, Part I*. New York: Wiley, 1968.

[11] P. Huber, *Robust Statistics*. Wiley-Interscience, 1981.

[12] Y. Wu and S. Verdú, "Rényi Information Dimension: Fundamental Limits of Almost Lossless Analog Compression," submitted to *IEEE Transactions on Information Theory*, 2009.

[13] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Proceedings of the Forty-Seventh Annual Allerton Conference on Communication, Control, and Computing*, Monticello, IL, October 2009.

[14] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proceedings of the National Academy of Sciences*, vol. 106, no. 45, pp. 18 914 – 18 919, November 2009.

[15] K. Falconer, *Techniques in Fractal Geometry*. John Wiley, 1997.