Scanning, Filtering and Prediction for Random Fields Corrupted by Gaussian Noise

Asaf Cohen
Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, ISRAEL
Email: soofsoof@tx.technion.ac.il

Neri Merhav
Department of Electrical Engineering
Technion - Israel Institute of Technology
Haifa 32000, ISRAEL
Email: merhav@ee.technion.ac.il

Tsachy Weissman
Department of Electrical Engineering
Stanford University
Stanford 94305-9510, USA
Email: tsachy@stanford.edu

***Abstract*—We consider the problem of sequential decision making on random fields corrupted by Additive White Gaussian Noise (AWGN). In particular, we first consider the problem of sequentially filtering an AWGN-corrupted random field. In this scenario, the sequential filter may be given the freedom to choose the path over which it traverses the random field (e.g., noisy image), thus it is natural to ask what is the best achievable performance and how far is the performance of widely used scanning methods from the optimum. We formally define the problem of scanning and filtering, derive a bound on the best achievable performance and quantify the excess loss occurring when non-optimal scanners are used, compared to optimal scanning and filtering.**

We then discuss the problem of sequential scanning and prediction of noisy random fields. This setting is a natural model for applications such as restoration and coding of noisy images. In this scenario, using predictive coding methods on the noisy image results in both enhancement and compression of the input image, as one expects that the prediction error consists mainly of the noise signal. We formally define the problem of sequential prediction in a noisy array and compute the optimal performance in terms of the clean scandictability defined by Merhav and Weissman.

I. Introduction

Consider the problem of sequentially scanning and filtering (or predicting) a multidimensional noisy data array, while minimizing a given loss function. Particularly, at each time instant t, $1 \leq t \leq |B|$, where $|B|$ is the number of sites ("pixels") in the data array, the sequential decision maker chooses a site to be visited, denoted by Ψ_t. In the filtering scenario, it first observes the value at that site, and then gives an estimation for the underlying clean value. In the prediction scenario, it is required to give a prediction for that (clean) value, before the actual observation is made. In both cases, both the location Ψ_t and the estimation or prediction may depend on the previously observed values - the values at sites Ψ_1 to Ψ_{t-1}. The goal is to minimize the cumulative loss after scanning the entire data array.

Important applications of this problem can be found in image and video processing, such as filtering or predictive coding. In these applications, one wishes to either enhance or jointly enhance and code a given image. The motivation behind a prediction/compression-based approach is that the prediction error may consist mainly of the noise signal, while the clean signal is recovered by the predictor. For example, see [1]. It is clear that different scanning patterns of the image may result in different filtering or prediction errors, thus, it is natural to ask what is the performance of the optimal scanning strategy, and what is the loss when non-optimal strategies are used.

An information theoretic discussion of the scanning problem was initiated by Lempel and Ziv in [2], where the Peano-Hilbert scan was shown to be optimal for compression of individual images. In [3], Merhav and Weissman formally defined the "scandictability" of a random field, namely, the best achievable performance for scanning and prediction of a random field, and discussed particular cases where this value can be computed and the optimal scanning order can be identified. In [4], a *universal* scanning and prediction algorithm which achieves the scandictability of any stationary random filed was given. Furthermore, the scenario of binary fields corrupted by binary noise was considered, and bounds on the best achievable performance and excess loss when non-optimal scanners are used were given. A more comprehensive survey can be found in [5] and [6].

In this paper, we extend some of the above results to the case where the decision maker observes only a noisy observation of the data, yet it is judged with respect to the clean data. In Section II, we give a precise definition of the problem. In Section III, we first give a bound on the best achievable performance of a scanner-filter pair. We then present the main contribution of this paper, which is the derivation of excess loss bounds for the case where non-optimal scanning strategies are used. That is, we assume that, due to implementation constraints, for example, one cannot use the optimal scanner for a given data array, and is forced to use a standard scanning order (e.g., raster scan). We focus our attention on Gaussian and binary input fields (corrupted by additive white Gaussian noise), and derive upper bounds on the excess loss incurred, compared to optimal scanning and filtering. In Section IV, we consider the prediction scenario, and characterize the best achievable performance in terms of the best achievable performance for the clean data prediction problem [3]. The special case of Gaussian fields corrupted by additive white Gaussian noise is discussed thoroughly. Further results on the filtering and prediction of random fields can be found in [6], where we also discuss the discrete case and universal algorithms.

II. Problem Formulation

We start with a formal definition of the problem. Let A denote the alphabet, which is either discrete or the real line. Let N be the noisy observation alphabet. Let $\Omega = (A \times N)^{\mathbb{Z}^2}$ be the observation space (the results can be extended to any finite dimension). A probability measure Q on Ω is stationary if it is invariant under translations τ_i, where for each $\omega \in \Omega$ and $i, j \in \mathbb{Z}^2$, $\tau_i(\omega)_j = \omega_{j+i}$ (namely, stationarity means shift invariance). Denote by $\mathcal{M}(\Omega)$ and $\mathcal{M}_S(\Omega)$ the sets of all probability measures on Ω and stationary probability measures on Ω, respectively. In this paper, we assume that $Q \in \mathcal{M}_S(\Omega)$, the probability measure governing the clean and noisy data, is known. Elements of $\mathcal{M}(\Omega)$, *random fields*, will be denoted by upper case letters while elements of Ω, *individual data arrays*, will be denoted by the corresponding lower case. It will also be beneficial to refer to the clean and noisy random fields separately, that is, $\{X_t\}_{t \in \mathbb{Z}^2}$ represents the clean signal and $\{Y_t\}_{t \in \mathbb{Z}^2}$ represents the noisy observations.

Let $\mathcal{V}$ denote the set of all finite subsets of $\mathbb{Z}^2$. For $V \in \mathcal{V}$, denote by X_V the restrictions of the data array X to V. For $i \in \mathbb{Z}^2$, X_i is the random variable corresponding to X at site i. Let $\mathcal{R}_\square$ be the set of all rectangles of the form $V = \mathbb{Z}^2 \cap ([m_1, m_2] \times [n_1, n_2])$. As a special case, denote by V_n the square $\{0, \ldots, n-1\} \times \{0, \ldots, n-1\}$. For $V \subset \mathbb{Z}^2$, let the interior radius of V be

$$R(V) \stackrel{\triangle}{=} \sup\{r : \exists c \text{ s.t. } B(c, r) \subseteq V\}, \tag{1}$$

where $B(c, r)$ is a closed ball (under the l_1-norm) of radius r centered at c. Throughout, $\log(\cdot)$ will denote the natural logarithm, and entropies will be measured in nats.

Definition 1: A *scanner-filter pair* for a finite set of sites $B \in \mathcal{V}$ is the following pair (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$ is a sequence of measurable mappings, $\Psi_t : N^{t-1} \mapsto B$ determining the site to be visited at time t, with the property that
$$\Big\{\Psi_1, \Psi_2(y_{\Psi_1}), \Psi_3(y_{\Psi_1}, y_{\Psi_2}) \ldots, \Psi_{|B|}\left(y_{\Psi_1}, \ldots, y_{\Psi_{|B|-1}}\right)\Big\} = B, \quad \forall y \in N^B. \tag{2}$$
- $\{\tilde{F}_t\}_{t=1}^{|B|}$ is a sequence of measurable functions, where for each t, $\tilde{F}_t : N^t \mapsto D$ determines the reconstruction for value at the site visited at time t, based on the current and previous observations, and D is the reconstruction alphabet.

Note that both the scanner Ψ and the filter $\{\tilde{F}_t\}$ base their decisions only on the noisy observations. The cumulative loss of a scanner-filter pair $(\Psi, \tilde{F})$ up to time $t \le |B|$ is denoted by $L_{(\Psi,\tilde{F})}(x_B, y_B)_t$,

$$L_{(\Psi,\tilde{F})}(x_B, y_B)_t = \sum_{i=1}^{t} l(x_{\Psi_i}, \tilde{F}_i(y_{\Psi_1}, \ldots, y_{\Psi_i})), \tag{3}$$

where $l : A \times D \mapsto [0, \infty)$ is the loss function. The sum of the instantaneous losses over the entire data array B, $L_{(\Psi,\tilde{F})}(x_B, y_B)_{|B|}$, will be abbreviated as $L_{(\Psi,\tilde{F})}(x_B, y_B)$.

For a given loss function l and field $Q \in \mathcal{M}(\Omega)$ restricted to B, define the best achievable scanning and filtering performance by

$$\tilde{U}(l, Q_B) = \inf_{(\Psi,\tilde{F}) \in \mathcal{S}(B)} E_{Q_B} \frac{1}{|B|} L_{(\Psi,\tilde{F})}(X_B, Y_B), \tag{4}$$

where Q_B is the marginal probability measure restricted to B and $\mathcal{S}(B)$ is the set of *all* possible scanner-filter pairs for B. The best achievable scanning and filtering performance for the field Q, $\tilde{U}(l, Q)$, is defined by

$$\tilde{U}(l, Q) = \lim_{n \to \infty} \tilde{U}(l, Q_{V_n}), \tag{5}$$

if this limit exists.

In the prediction scenario, $\bar{F}_t : N^{t-1} \mapsto D$ is allowed to base its estimation only on $y_{\Psi_1}, \ldots, y_{\Psi_{t-1}}$, and we have

$$L_{(\Psi,\bar{F})}(x_B, y_B) = \sum_{t=1}^{|B|} l(x_{\Psi_t}, \bar{F}_t(y_{\Psi_1}, \ldots, y_{\Psi_{t-1}})), \tag{6}$$

$$\bar{U}(l, Q_B) = \inf_{(\Psi,\bar{F})} E_{Q_B} \frac{1}{|B|} L_{(\Psi,\bar{F})}(X_B, Y_B), \tag{7}$$

and

$$\bar{U}(l, Q) = \lim_{n \to \infty} \bar{U}(l, Q_{V_n}), \tag{8}$$

if this limit exists.

The following proposition asserts, analogously to [3, Theorem 1], that for any stationary random field the limits (5) and (8) both exist.

Proposition 2: For any stationary field $Q \in \mathcal{M}_S(\Omega)$ and for any sequence $\{B_n\}$, $B_n \in \mathcal{R}_\square$, satisfying $R(B_n) \to \infty$, the limits in (5) and (8) exist and satisfy

$$\tilde{U}(l, Q) = \lim_{n \to \infty} \tilde{U}(l, Q_{B_n}) = \inf_{\Delta \in \mathcal{R}_\square} \tilde{U}(l, Q_\Delta), \tag{9}$$

$$\bar{U}(l, Q) = \lim_{n \to \infty} \bar{U}(l, Q_{B_n}) = \inf_{\Delta \in \mathcal{R}_\square} \bar{U}(l, Q_\Delta). \tag{10}$$

Since $\tilde{U}(l, Q_B)$ and $\bar{U}(l, Q_B)$, possess the sub-additivity property, e.g., for any $V, V', V \cap V' = \emptyset$, there exists a scanner-filter pair $(\Psi, \tilde{F})$ on $V \cup V'$ such that

$$E_Q L_{(\Psi,\tilde{F})}(X_{V \cup V'}, Y_{V \cup V'}) \le |V| \tilde{U}(l, Q_V) + |V'| \tilde{U}(l, Q_{V'}), \tag{11}$$

the proof of Proposition 2 follows verbatim from the proof of [3, Theorem 1].

III. Scanning and Filtering of AWGN-Corrupted Random Fields

In this section, we first give a bound on the best achievable scanning and filtering performance. Since the optimal performance may be difficult to achieve in practical situations, mainly due to the difficulty in identifying the optimal scan, we derive bounds on the sensitivity of the scanning and filtering performance to the scan.

We assume an invertible memoryless channel, meaning the input distribution of a single symbol is uniquely determined given the output distribution. As a simple example, a discrete memoryless channel with an invertible channel matrix can be kept in mind, however, the result below applies to more general channels, including continuous ones, and the Gaussian channel in particular. In this case, the associated Bayes envelope is

$$\phi_l(P) = \min_{f(\cdot)} El(X, f(Y)), \tag{12}$$

where P is the distribution of the channel output Y. Define

$$\zeta(d) = \max\{H(P) : \phi_l(P) \leq d\}, \tag{13}$$

where $H(P)$ is the differential entropy of a random variable distributed according to P, and let $\bar{\zeta}(\cdot)$ be the upper concave envelope of $\zeta(\cdot)$.

Theorem 3: Let Y_B be the output of an invertible memoryless channel whose input is X_B. Then, for any scanner-filter pair $(\Psi, \tilde{F})$ we have

$$\bar{\zeta}\left(\frac{1}{|B|} E_{Q_B} L_{\Psi,\tilde{F}}(X_B, Y_B)\right) \geq \frac{1}{|B|} H(Y_B), \tag{14}$$

where $H(Y_B)$ is the differential entropy of Y_B. That is,

$$\bar{\zeta}\left(\tilde{U}(l, Q_B)\right) \geq \frac{1}{|B|} H(Y_B). \tag{15}$$

Theorem 3 is the direct analogue of the lower bounds in [3] for the filtering scenario. Notice that $\bar{\zeta}$ is a single-letter function, whose value depends only on the scalar filtering problem. The dependencies in the data are taken into account only through $H(Y_B)$. $\bar{\zeta}$, though, is not easily computed [6]. Finally, note that Theorem 3 holds for any finite B. A similar result can be derived for the clean data prediction problem, strengthening the asymptotic results of [3].

A. Bounds on the Excess Loss of Non Optimal Scanners

Theorem 3 gives a lower bound on best possible scanning and filtering performance. It is interesting to investigate, from both practical and theoretical reasons, what is the excess scandiction loss when non-optimal scanners are used. Such results can shed light on the performance of different scanning patterns such as the Peano scan, or the scans considered by Memon *et. al.* in [7].

The results herein are achieved by bounding the absolute difference between the scanning and filtering performance of any two scans, Ψ^1 and Ψ^2, assuming both use their optimal filters. These bounds result from a relation between the performance of *discrete time* filtering and *continuous time* filtering, together with the fundamental results of Duncan [8] and [9] on the relation between mutual information and causal minimum mean square error estimation. From now on we assume the loss function is the squared error loss, denoted l_s.

We start with several definitions. With a slight abuse of notations, let X be a Gaussian random variable, $X \sim \mathcal{N}(0, \sigma_X^2)$. Consider the following two estimation problems:

- Estimating X based on $Y = X + N$, where $N \sim \mathcal{N}(0, \sigma_N^2)$, independent of X.
- Causally estimating $X_t \equiv X$, $t \in [0,1]$, based on Y_t, which is an AWGN-corrupted version of X_t, the Gaussian noise having a spectral density level of σ_N^2.

Note that since the integral over Y_t in the second estimation problem is a sufficient statistic in order to estimate X, the noise value in both problems is effectively the same. To bound the sensitivity of the scanning and filtering performance, we consider the difference

$$\int_0^1 \text{Var}(X|Y^t)dt - \text{Var}(X|Y). \tag{16}$$

It is not hard to show [6] that

$$\int_0^1 \text{Var}(X|Y^t)dt - \text{Var}(X|Y) = \sigma_N^2 f\left(\frac{\sigma_X^2}{\sigma_N^2}\right), \tag{17}$$

where

$$f(x) = \ln(1+x) - \frac{x}{x+1}. \tag{18}$$

In words, the expression in the right hand side of (17) is simply the difference between the performances of two restoration problems: a continuous-time filtering problem and a scalar (hence, discrete) estimation problem. The following theorem is the main result in this paper. It bounds the absolute difference between the performance of *any* two scanners, assuming each scanner is equipped with its optimal filter, thus assessing the sensitivity of the performance to the scan itself.

Theorem 4: Let X_{V_n} be a Gaussian random field with a constant marginal distribution satisfying $\text{Var}(X_i) = \sigma_X^2 < \infty$ for all $i \in V_n$. Let $Y_i = X_i + N_i$, where N_{V_n} is a white Gaussian noise of variance σ_N^2, independent of X_{V_n}. Then, for any two scans Ψ^1 and Ψ^2, we have

$$\frac{1}{n^2}\left|EL_{(\Psi^1,\tilde{F}^{opt})}(X_{V_n}, Y_{V_n}) - EL_{(\Psi^2,\tilde{F}^{opt})}(X_{V_n}, Y_{V_n})\right| \leq \sigma_N^2 f\left(\frac{\sigma_X^2}{\sigma_N^2}\right). \tag{19}$$

At this point, a few remarks are in order. The bound in Theorem 4 is applicable only to Gaussian random fields corrupted by AWGN. A very simple bound, applicable to arbitrarily distributed fields results from noting that for *any* random variable X,

$$0 \leq \text{Var}(X|Y) \leq \sigma_N^2 \frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2}. \tag{20}$$

Namely, simple symbol by symbol restoration results in a cumulative loss of $\sigma_N^2 \frac{\sigma_X^2}{\sigma_X^2+\sigma_N^2}$, so the excess loss in non-optimal scanning cannot be greater than that value. Hence,

$$\frac{1}{n^2}\left|EL_{(\Psi_1,\tilde{F}^{opt})}(X_{V_n}, Y_{V_n}) - EL_{(\Psi_2,\tilde{F}^{opt})}(X_{V_n}, Y_{V_n})\right| \leq \sigma_N^2 \frac{\sigma_X^2}{\sigma_X^2 + \sigma_N^2}. \tag{21}$$

Both the bound in Theorem 4 and the bound in (21) are in the form of $\text{Var}(X_1)g(\text{SNR})$, for some g, where $\text{SNR} = \sigma_X^2/\sigma_N^2$. This means that any bound obtained for a certain SNR applies to all values of $\text{Var}(X_1)$ by rescaling. The bound in Theorem 4 has the form $\text{Var}(X_1)\frac{f(\text{SNR})}{\text{SNR}}$, and we have

$$\lim_{\text{SNR}\to 0^+} \frac{f(\text{SNR})}{\text{SNR}} = \lim_{\text{SNR}\to\infty} \frac{f(\text{SNR})}{\text{SNR}} = 0, \tag{22}$$

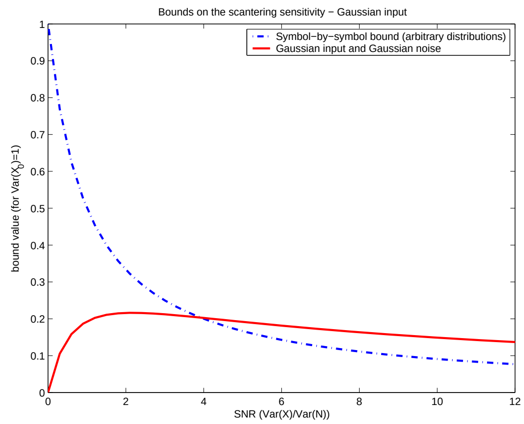

Fig. 1. Bounds on the excess loss in scanning and filtering of arbitrary input (dashed) and Gaussian input (solid) under AWGN.

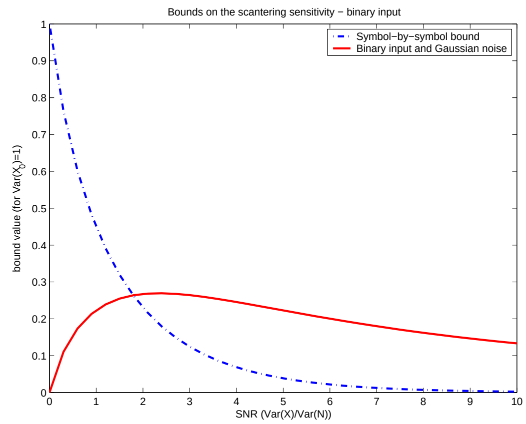

Fig. 2. Bounds on the excess loss in scanning and filtering of binary input fields corrupted by AWGN. The solid line is the bound given in Theorem 5, and the dashed line is the bound given in (26).

that is, the scan is inconsequential at very high or very low SNR. This is clear as at high SNR the current observation is by far the most influential, and whatever previous observations used is inconsequential. For low SNR, the cumulative loss is high whatever the scan is. Unlike the bound in Theorem 4, (21) does not capture the correct behavior for SNR $\to 0^+$, and is mainly interesting in the high SNR regime.

The above observations are also evident in Fig. 1, which includes both the bound given in Theorem 4, applicable to Gaussian input fields, and (21), applicable to arbitrary input fields. It is also evident that in the case of a Gaussian input sequence, $\frac{f(\text{SNR})}{\text{SNR}}$ has a unique maximum of approximately 0.216, that is, the excess loss due to a suboptimal scan at any SNR is upper bounded by $0.216\text{Var}(X_1)$.

The bound in Theorem 4 can be generalized to non-Gaussian inputs [6]. In the case where X_{V_n} is a binary random field, with a symmetric marginal distribution (that is, $P(X_0 = \sigma_X) = P(X_0 = -\sigma_X) = 1/2$, but the X_i's are not necessarily i.i.d., and any dependence between them is possible) and Y_{V_n} is the AWGN-corrupted version of X_{V_n}, the bound admits the following simple form.

Theorem 5: Let X_{V_n} be a binary random field with a constant symmetric marginal distribution satisfying $\text{Var}(X_i) = \sigma_X^2 < \infty$ for all $i \in V_n$. Let $Y_i = X_i + N_i$, where N_{V_n} is a white Gaussian noise of variance σ_N^2, independent of X_{V_n}. Then, for any two scans Ψ^1 and Ψ^2, we have

$$\frac{1}{n^2}\left|EL_{(\Psi^1,\tilde{F}^{opt})}(X_{V_n},Y_{V_n}) - EL_{(\Psi^2,\tilde{F}^{opt})}(X_{V_n},Y_{V_n})\right| \leq 2\sigma_N^2 \text{I}(\text{SNR}) - \sigma_X^2 \text{mmse}(\text{SNR}), \quad (23)$$

where

$$\text{I}(\text{SNR}) = \text{SNR} - \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \log\cosh(\text{SNR} - \sqrt{\text{SNR}}y)dy, \quad (24)$$

and

$$\text{mmse}(\text{SNR}) = 1 - \int_{-\infty}^{\infty} \frac{e^{-\frac{y^2}{2}}}{\sqrt{2\pi}} \tanh(\text{SNR} - \sqrt{\text{SNR}}y)dy. \quad (25)$$

The above bound is plotted in Fig. 2. Similarly as in the case of Gaussian input, it is insightful to compare this bound to a simple symbol-by-symbol filtering bound. That is, since for any binary X corrupted by AWGN of variance σ_N^2, $0 \leq \text{Var}(X|Y) \leq \sigma_X^2 \text{mmse}(\text{SNR})$, we have

$$\frac{1}{n^2}\left|EL_{(\Psi^1,\tilde{F}^{opt})}(X_{V_n},Y_{V_n}) - EL_{(\Psi^2,\tilde{F}^{opt})}(X_{V_n},Y_{V_n})\right| \leq \sigma_X^2 \text{mmse}(\text{SNR}), \quad (26)$$

where mmse(SNR) is given in (25). This is simply the analogue of (21) to the binary input setting.

The bounds in Theorems 4 and 5 are tight for low SNR, but much losser at high SNR. This is clear, as at low SNR the difference between the continuous-time and discrete-time estimation problems vanishes, and this difference is at the heart of the bounding technique. At high SNR, since the current observation is the most influential, even simple symbol-by-symbol bounds such as (21) or (26) give meaningfull results.

IV. Sequential Prediction Under Squared Error

So far, we discussed the filtering problem. We now turn to consider the prediction scenario. Let Q_Y denote the marginal distribution of the noisy observations field Y. Let $U(l, Q_Y)$ denote the "clean" scandictability of Y, as defined in [3]. That is, at each time t one wishes to predict Y_{Ψ_t} based on $Y_{\Psi_1}^{\Psi_{t-1}}$. In this section, we describe the noisy scandictability for the additive noise model and the squared error loss function, in terms of the clean scandictability of Y, and give the optimal scandictor. The following lemma is proved in [6].

Lemma 6: Let $\{(X_t, Y_t)\}_{t\in\mathbb{Z}^2}$ be a random field governed by a probability measure Q such that $Y_t = X_t + N_t$, where N_t, $t \in \mathbb{Z}^2$, are i.i.d. random variables with $\text{Var}(N_t) = \sigma_N^2 < \infty$. Let $\{B_n\}_{n\geq 1}$ be any sequence of elements in $\mathcal{V}$, satisfying $R(B_n) \to \infty$. Then

$$\bar{U}(l_s, Q) = U(l_s, Q_Y) - \sigma_N^2. \quad (27)$$

Furthermore, $\bar{U}(l_s, Q)$ is achieved by the scandictor which achieves $U(l_s, Q_Y)$.

Actually, Lemma 6 is only scarcely related to scanning. It merely states that in the prediction of a process based on its noisy observations, under the additive model stated above and squared error loss, the optimal predictor is a one which disregards the noise, and attempts to predict the next *noisy* outcome. Similar results for binary processes through a BSC where given in [10]. Note that using Lemma 6, one can derive bounds on the best achievable performance $\bar{U}$ using the bounds on U derived in [3].

A. Gaussian Fields

Let both X and N be Gaussian random fields, where the components of N are i.i.d. and independent of X. That is, Y is the output of an additive white Gaussian noise channel, with a Gaussian input X. In this scenario, similarly to the clean one, the noisy scandictability is known exactly and is given by a single letter expression.

Before we proceed, several definitions are required. For any $t \in \mathbb{Z}^2$ and $V \subseteq \mathbb{Z}^2$, denote by $\hat{X}_t(V)$ the best linear predictor of X_t given $\{X_{t'}\}_{t' \in V}$. A subset $S \subseteq \mathbb{Z}^2$ is called a *half plane* if it is closed to addition and satisfies $S \cup (-S) = \mathbb{Z}^2$ and $S \cap (-S) = \{0\}$. For example, $S_{\text{lex}} = \{(m,n) \in \mathbb{Z}^2 : [m > 0] \text{ or } [m = 0, n \geq 0]\}$ is a half plane. Let X be a wide sense stationary random field and denote by g the density function associated with the absolutely continuous component in the Lebesgue decomposition of its spectral measure. Then, for any half plane S, we have [11, Theorem 1],

$$E\big(X_0 - \hat{X}_0(-S \setminus \{0\})\big)^2 = \exp\left\{\frac{1}{4\pi^2}\int_{[0,2\pi)^2} \log g(\lambda) d\lambda\right\}, \tag{28}$$

denoted $\sigma_u^2(X)$. We can now state the following corollary, regarding the noisy scandictability in the Gaussian regime and squared error loss, which is a trivial application of Lemma 6 and the results of [3, Section IV].

Corollary 7: Under the terms of Lemma 6, assuming both X and N are Gaussian, the noisy scandictability of Q is given by

$$\bar{U}(l_s, Q) = \sigma_u^2(Y) - \sigma_N^2. \tag{29}$$

Furthermore, $\bar{U}(l_s, Q)$ is asymptotically achieved by a scandictor which scans (X_t, Y_t) according to the total order defined by any half-plane S and applies the corresponding best linear predictor for the next outcome of Y.

For any stationary Gaussian *process* X, it has been shown by Kolmogorov (e.g., [12]) that the entropy rate is given by

$$H_*^X = \frac{1}{2}\log(2\pi e) + \frac{1}{4\pi}\int_{-\pi}^{\pi} \log g(\lambda) d\lambda. \tag{30}$$

Thus, $H_*^X = \frac{1}{2}\log(2\pi e \sigma_u^2(X))$, and we have,

$$\bar{U}(l_s, Q) = \frac{1}{2\pi e} e^{2H_*^Y} - \frac{1}{2\pi e} e^{2H^N}, \tag{31}$$

where H_*^Y is the entropy rate of Y and H^N is the entropy of each N_t. From the entropy power inequality, we have,

$$e^{2H_*^Y} \geq e^{2H^N} + e^{2H_*^X}, \tag{32}$$

thus, as expected, the noisy scandictability given in Corollary 7 (and (31)) is at least as large as the clean scandictability of X, that is, with no noise at all. In most of the interesting cases, however, (32) is a strict inequality. In fact (32) is achieved with equality only when both X and N are Gaussian and have *proportional spectra*. Consequently, unless X is white noise, Corollary 7 is non-trivial.

Finally, we mention that excess loss bounds in the spirit of Section III-A can be derived for the prediction scenario as well, using methods similar to those used in Theorems 4 and 5 or new ones. The details are in [6].

V. Conclusion

We investigated problems in sequential filtering and prediction of multidimensional data arrays. A special emphasis was given to the case of AWGN and squared error loss. A bound on the best achievable scanning and filtering performance was derived, and the excess loss incurred when non-optimal scanners are used was quantified. In the prediction setting, a relation of the best achievable performance to that of the clean scandictability was given.

Although the problem is strongly related to its one-dimensional analogue, the numerous scanning possibilities result in a richer and more challenging problem. Moreover, while results for prediction of noisy data arrays can be deduced from the clean analogue using modified loss functions, the filtering scenario required new tools and techniques. Future work may aim at identifying optimal scanning and filtering methods and deriving new bounds for different channel models.

References

[1] B. Natarajan, K. Konstantinides, and C. Herley, "Occam filters for stochastic sources with application to digital images," *IEEE Trans. Signal Processing*, vol. 46, no. 5, pp. 1434–1438, May 1998.

[2] A. Lempel and J. Ziv, "Compression of two-dimensional data," *IEEE Trans. Inform. Theory*, vol. IT-32, no. 1, pp. 2–8, January 1986.

[3] N. Merhav and T. Weissman, "Scanning and prediction in multidimensional data arrays," *IEEE Trans. Inform. Theory*, vol. 49, no. 1, pp. 65–82, January 2003.

[4] A. Cohen, N. Merhav, and T. Weissman, "Universal scanning and sequential decision making for multi-dimensional data," in *Proc. Int. Symp. Inf. Th.*, Seattle, USA, July 2006, pp. 431–435.

[5] A. Cohen, N. Merhav, and T. Weissman, "Universal scanning and sequential decision making for multi-dimensional data - part I: the noiseless case," submited to *IEEE Trans. on Inform. Theory*, September 2006.

[6] A. Cohen, T. Weissman, and N. Merhav, "Scanning and sequential decision making for multi-dimensional data - part II: the noisy case," 2007, in preparation.

[7] N. D. Memon, K. Sayood, and S. S. Magliveras, "Lossless image compression with a codebook of block scans," *IEEE Journal on Selected Areas In Communications*, vol. 13, no. 1, pp. 24–30, January 1995.

[8] T. E. Duncan, "On calculation of mutual information," *SIAM Journal of Applied Mathematics*, vol. 19, pp. 215–220, July 1970.

[9] D. Guo, S. Shamai, and S. Verdú, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inform. Theory*, vol. 51, no. 4, pp. 1261–1282, April 2005.

[10] T. Weissman and N. Merhav, "Universal prediction of random binary sequences in a noisy environment," *Ann. App. Prob.*, vol. 14, no. 1, pp. 54–89, February 2004.

[11] H. Helson and D Lowdenslager, "Prediction theory and Fourier series in several variables," *Acta Math.*, vol. 99, pp. 165–202, 1958.

[12] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 2nd edition, 1984.