Universal Scanning and Sequential Decision Making for Multidimensional Data

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 investigate several problems in scanning of multidimensional data arrays, such as universal scanning and prediction ("scandiction", for short), and scandiction of noisy data arrays. These problems arise in several aspects of image and video processing, such as predictive coding, filtering and denoising. In predictive coding of images, for example, an image is compressed by coding the prediction error sequence resulting from scandicting it. Thus, it is natural to ask what is the optimal method to scan and predict a given image, what is the resulting minimum prediction loss, and if there exist specific scandiction schemes which are universal in some sense.**

More specifically, we investigate the following problems: First, given a random field, we examine whether there exists a scandiction scheme which is independent of the field's distribution, yet asymptotically achieves the same performance as if this distribution was known. This question is answered in the affirmative for the set of all spatially stationary random fields and under mild conditions on the loss function. We then discuss the scenario where a non-optimal scanning order is used, yet accompanied by an optimal predictor, and derive a bound on the excess loss compared to optimal scandiction. Finally, we examine the scenario where the random field is corrupted by noise, but the scanning and prediction (or filtering) scheme is judged with respect to the underlying noiseless field.

I. Introduction

The problem of sequentially predicting the next outcome of a sequence, x_t, based on the previously observed outcomes, $x_1, x_2, \ldots, x_{t-1}$, is well-studied. The current literature includes numerous results regarding the very many settings in which this problem can be examined. The problem of prediction in multidimensional data arrays, however, has received far less attention so far. In this scenario, apart from the on-line strategies for the sequential prediction of the data, the fundamental problem of *scanning* it should be considered. The problem mainly arises in image compression, where various methods of predictive coding are used (e.g., [1]). In this case, the encoder may be given the freedom to choose the actual *path* over which it traverses the image, and thus it is natural to ask which path is optimal in the sense of minimal cumulative prediction loss (which may result in maximal compression).

In [2], a specific scanning method was suggested by Lempel and Ziv for the lossless compression of multidimensional data. It was shown that the application of the incremental parsing algorithm of [3] on the one dimensional sequence resulting from the *Peano-Hilbert* scan yields a universal compression algorithm with respect to all finite-state *scanning and encoding*. These results where later extended in [4] to the probabilistic setting, where it was shown that this algorithm is also universal for any stationary Markov random field. Furthermore, Using a universal quantization algorithm, the existence of a universal rate-distortion encoder was also established. Lossless image compression was also discussed in [5], where it was shown that in the special case of first order predictors (i.e., with context of length one), the optimal scanning order can be viewed as a minimal spanning tree of a graph, whose vertices are the image pixels. However, since the number of bits required to encode the resulting scanning order may be prohibitively high, the authors suggested a sub-optimal method which used a *codebook* of scans.

The results of [2] and [4] considered a specific, data independent scan of the data set. However, for a general loss function and random field (or individual image), it is not clear that the Peano-Hilbert scan will remain optimal. This more general scenario was discussed in [6], where Merhav and Weissman formally defined the notion of a *scandictor*, a scheme for both scanning and prediction, as well as that of *scandictability*, the best expected performance on a data array. The main result in [6] is the fact that if a stochastic field can be represented autoregressively (under a specific scan Ψ) with a maximum-entropy innovation process, then it is optimally scandicted in the way it was created (i.e., by the specific scan Ψ and its corresponding optimal predictor).

While defining the yardstick for analyzing scanning and prediction in multidimensional arrays, the work in [6] leaves many open challenges. As the topic of prediction in one-dimensional arrays is rich and includes elegant solutions to various prediction problems, seeking analogous results in the multidimensional case offers plentiful research objectives.

In Section III, we consider the case where one strives to compete with a finite set $\mathcal{F}$ of scandictors. Specifically, assume that there exists a probability measure Q which governs the data array. Of course, given the probability measure Q and the scandictor set, one can compute the optimal scandictor in the set (in some sense which will be defined later). However, we are interested in a universal scandictor, which scans the data independently of Q, and yet achieves essentially the same results as the best scandictor in $\mathcal{F}$. We show that it is indeed possible to compete with any finite set of scandictors when the random field is spatially stationary, and use this result to show that there exists a universal scandictor which achieves the scandictability of any spatially stationary random field.

In Section IV, we introduce an upper bound on the excess loss incurred when non-optimal scanners are used, yet with optimal prediction schemes. Namely, we consider the scenario where one cannot use a universal scandictor (or the optimal scan for a given random field), and instead uses an arbitrary scanning order, accompanied by the optimal predictor for that scan.

In [7] (and references therein), Weissman and Merhav extended the problem of universal prediction to the case of a *noisy* environment. Namely, the predictor observes a noisy version of the sequence, yet, it is judged with respect to the clean sequence. The problems of prediction and filtering in a noisy environment are intimately related to that of *denoising*, the latter having numerous applications, in several fields such as communications, image processing and more. In Section V, we discuss the case of scanning and *filtering* a noisy data array ("scantering", for short), give a lower bound on the best possible scantering performance, and bound the excess loss incurred when a non-optimal scanner is used in the binary filtering problem.

II. Problem Formulation

The following notation will be used throughout this paper. Let A denote the alphabet, which is either discrete or the real line. Let $\Omega = A^{\mathbb{Z}^d}$ denote the space of all possible data arrays in $\mathbb{Z}^d$. Although the results in this paper are applicable to any $d \geq 1$, for simplicity, they are formulated for $d = 2$. A probability measure Q on Ω is stationary if it is invariant under translations τ_i, $i \in \mathbb{Z}^2$ (i.e., shift invariant). Denote by $\mathcal{M}(\Omega)$ and $\mathcal{M}_S(\Omega)$ the spaces of all probability measures and stationary probability measures on Ω, respectively. 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.

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. Denote by V_n the square $\{0, \ldots, n-1\} \times \{0, \ldots, n-1\}$.

Definition 1 ([6]): A *scandictor* for the 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, $\Psi_t : A^{t-1} \mapsto B$ determining the site to be visited at time t, with the property that
$$\begin{aligned}\{\Psi_1, \Psi_2(x_{\Psi_1}), \Psi_3(x_{\Psi_1}, x_{\Psi_2}), \ldots \\ \Psi_{|B|}\left(x_{\Psi_1}, \ldots, x_{\Psi_{|B|-1}}\right)\} = B, \quad \forall x \in A^B. \quad (1)\end{aligned}$$
- $\{F_t\}_{t=1}^{|B|}$ is a sequence of measurable predictors, $F_t : A^{t-1} \mapsto D$ determining the prediction for the site visited at time t based on the observations at past visited sites, where D is the prediction alphabet.

We allow *randomized scandictors*, namely, scandictors such that $\{\Psi_t\}_{t=1}^{|B|}$ or $\{F_t\}_{t=1}^{|B|}$ can be chosen randomly from some set of possible functions. Note that scandictors for *infinite* data arrays are not considered in this paper. Definition 1, and the results to follow, consider only scandictors for finite sets of sites, ones which can be viewed merely as a reordering of the sites in a finite set B. We will consider, though, the limit as the size of the array tends to infinity.

Denote by $L_{(\Psi,F)}(x_{V_n})$ the cumulative loss of (Ψ, F) over x_{V_n}, that is

$$L_{(\Psi,F)}(x_{V_n}) = \sum_{t=1}^{|V_n|} l\left(x_{\Psi_t}, F_t(x_{\Psi_1}, \ldots, x_{\Psi_{t-1}})\right), \quad (2)$$

where $l : A \times D \to [0, \infty)$ is a given loss function. Throughout this paper, we assume that $l(\cdot, \cdot)$ is non-negative and bounded by $l_{max} < \infty$. The scandictability of a source $Q \in \mathcal{M}(\Omega)$ on $B \in \mathcal{V}$ is defined by

$$U(l, Q_B) = \inf_{(\Psi,F) \in \mathcal{S}(B)} E_{Q_B} \frac{1}{|B|} L_{(\Psi,F)}(X_B), \quad (3)$$

where Q_B is the marginal probability measure of X restricted to B and $\mathcal{S}(B)$ is the set of *all* possible scandictors for B. The scandictability of $Q \in \mathcal{M}(\Omega)$ is defined by

$$U(l, Q) = \lim_{n \to \infty} U(l, Q_{V_n}), \quad (4)$$

whenever the limit exists. By [6, Theorem 1], the limit in (4) exists for any $Q \in \mathcal{M}_S(\Omega)$.

It will be constructive to refer to the *finite set* scandictability as well. Let $\mathcal{F} = \{\mathcal{F}_n\}$ be a sequence of finite sets of scandictors, where for each n, $|\mathcal{F}_n| = \lambda < \infty$, and the scandictors in $\mathcal{F}_n$ are defined for the finite set of sites V_n. A possible scenario is one in which one has a set of "scandiction rules", each of which defines a unique scanner for each n, yet all these scanners comply with the same rule. In this case, $\mathcal{F} = \{\mathcal{F}_n\}$ can also be viewed as one finite set $\mathcal{F}$ which includes sequences of scandictors. We may also consider cases in which $|\mathcal{F}_n|$ increases with n (but finite for finite n). For $Q \in \mathcal{M}_S(\Omega)$ and $\mathcal{F} = \{\mathcal{F}_n\}$, we thus define the finite set scandictability of Q as the limit

$$U_{\mathcal{F}}(l, Q) \triangleq \lim_{n \to \infty} \min_{(\Psi,F) \in \mathcal{F}_n} E_{Q_{V_n}} \frac{1}{|V_n|} L_{(\Psi,F)}(X_{V_n}), \quad (5)$$

if it exists.

III. Universal Scandiction

At first sight, in order to compete successfully with a finite set of scandictors, one may try to use known algorithms for learning with expert advice, e.g., the *exponential weighting* algorithm suggested in [8] or the work which followed it. In this algorithm, each expert is assigned with a weight, according to its past performance. By decreasing the weight of poorly performing experts, hence preferring the ones proved to perform well thus far, one is able to compete with the best expert, having neither any *a priori* knowledge on the input sequence nor which expert will perform the best. However, in the scandiction problem, as each of the experts may use a different scanning strategy, at a given point in time each scanner might be at a different site, with different sites as its past. Thus, it is not at all guaranteed that one can alternate from one expert to the other. This obstacle is in someway similar to the one faced in [9], where the consideration of a loss function with memory prevented the alternation of experts each time instant, or to source coding problems such as [10]. The difficulties in these examples differ from those we confront here. Yet, the solution suggested therein, which is to persist on using the same expert for a significantly long block of data before alternating it, was found useful in our universal scanning problem.

Particularly, in order to establish the existence of a universal scandictor, we propose the following algorithm. Let x_{V_n} be the $n \times n$ data array to be scandicted. For $M < n$, define $K \triangleq \lceil \frac{n}{M} \rceil - 1$. Divide x_{V_n} into K^2 blocks of size $M \times M$ and $2K+1$ blocks of possibly smaller size. Denote by x^i, $0 \leq i \leq (K+1)^2 - 1$ the i'th block under some fixed scanning order of the blocks. This scanning order is irrelevant in this case, so we assume from now on that it is a (continuous) raster scan from the upper left corner. Let $\mathcal{F} = \{\mathcal{F}_n\}$ be the sequence of scandictor sets. The suggested algorithm scans the data in x_{V_n} block-wise, that is, it does not apply any of the scandictors in $\mathcal{F}_n$, only scandictors from $\mathcal{F}_M$. Omitting M for convenience, denote by $L_{j,i}$ the cumulative loss of $(\Psi, F)_j \in \mathcal{F}_M$ after scanning i blocks, where $(\Psi, F)_j$ is *restarted* after each block, namely, it scans each block separately and independently of the other blocks. Note that $L_{j,i} = \sum_{m=0}^{i-1} L_j(x^m)$ and that for $i = 0$, $L_{j,i} = 0$ for all j. Since we assumed the scandictors are capable of scanning only square blocks, for the $2K + 1$ possibly smaller (and not square) blocks the loss may be l_{max} throughout. For $\eta > 0$, and any i and j, define

$$P_i\left(j | \{L_{j,i}\}_{j=1}^{\lambda}\right) = \frac{e^{-\eta L_{j,i}}}{\sum_{j=1}^{\lambda} e^{-\eta L_{j,i}}}, \tag{6}$$

where $\lambda = |\mathcal{F}_M|$. For each $0 \leq i \leq (K+1)^2 - 1$, after scanning i blocks of data, the algorithm computes $P_i\left(j | \{L_{j,i}\}_{j=1}^{\lambda}\right)$ for each j. It then randomly selects a scandictor according to this distribution, independently of its previous selections, and uses this scandictor as its output for the $(i+1)$-st block.

The following theorem stands at the basis of our results, asserting the existence of a universal scandictor which competes successfully with any finite set of scandictors.

Theorem 2: Let X be a stationary random field with a probability measure Q. Let $\mathcal{F} = \{\mathcal{F}_n\}$ be an arbitrary sequence of scandictor sets, where $\mathcal{F}_n$ is a set of scandictors for V_n and $|\mathcal{F}_n| = \lambda < \infty$ for all n. Then, there exists a sequence of scandictors $(\hat{\Psi}, \hat{F})_n$, independent of Q, for which

$$\begin{aligned}\liminf_{n\to\infty} E_{Q_{V_n}} E \frac{1}{|V_n|} L_{(\hat{\Psi}, \hat{F})_n}(X_{V_n}) \\ \leq \liminf_{n\to\infty} \min_{(\Psi,F)\in\mathcal{F}_n} E_{Q_{V_n}} \frac{1}{|V_n|} L_{(\Psi,F)}(X_{V_n})\end{aligned} \tag{7}$$

for any $Q \in \mathcal{M}_S(\Omega)$, where the inner expectation in the l.h.s. of (7) is due to the possible randomization in $(\hat{\Psi}, \hat{F})_n$.

A detailed proof of Theorem 2, as well as proofs for the statements to follow, can be found in [11].

A. Finite-State Scandiction

Consider now the set of finite-state scandictors, very similar to the set of finite-state encoders described in [2]. At time $t = 1$, a *finite-state scandictor* starts at an arbitrary initial site Ψ_1, with an arbitrary initial state $s_0 \in S$ and gives $F(s_0)$ as its prediction for x_{Ψ_1}. Only then it observes x_{Ψ_1}. After observing x_{Ψ_i}, it computes its next state, s_i, according to $s_i = g(s_{i-1}, x_{\Psi_i})$ and advances to the next site, $x_{\Psi_{i+1}}$, according to $\Psi_{i+1} = \Psi_i + d(s_i)$, where $g : S \times A \mapsto S$ is the next state function and $d : S \mapsto B$ is the displacement function, $B \subset \mathbb{Z}^2$ denoting a fixed finite set of possible relative displacements. It then gives its prediction $F(s_i)$ to the value $x_{\Psi_{i+1}}$. Similarly to [2], we assume the alphabet A includes an additional "End of File" (EoF) symbol to mark the image edges. The following lemma and the theorem which follows establish the fact that the set of finite-state scandictors is indeed rich enough to achieve the scandictability of any stationary source, yet not too rich to compete with.

Lemma 3: Let $\mathcal{F}_\nu = \{(\Psi, F)_j\}$ be the set of all finite-state scandictors with at most ν states. Then, for any $Q \in \mathcal{M}_S(\Omega)$, $\lim_{\nu\to\infty} U_{\mathcal{F}_\nu}(l, Q) = U(l, Q)$. That is, the scandictability of any spatially stationary source is asymptotically achieved with finite-state scandictors.

Assume now that both the source alphabet A and the prediction alphabet D are finite. The following theorem asserts, under the above assumption, the existence of a universal scandictor for all stationary random fields.

Theorem 4: Let X be a stationary random field over a finite alphabet A and a probability measure Q. Assume that the prediction alphabet D is finite. Then, there exists a sequence of scandictors $(\Psi, F)_n$, independent of Q, for which

$$\lim_{n\to\infty} E_{Q_{V_n}} E \frac{1}{|V_n|} L_{(\Psi,F)_n}(X_{V_n}) = U(l, Q) \tag{8}$$

for any $Q \in \mathcal{M}_S(\Omega)$, where the inner expectation in the l.h.s. of (8) is due to the possible randomization in $(\Psi, F)_n$.

IV. Bounds on the Excess Scandiction Loss for Non-Optimal Scanners

While the results of Section III establish the existence of a universal scandictor for all stationary random fields and bounded loss function (under the terms of Theorem 4), it is interesting to investigate, from both practical and theoretical reasons, what is the excess scandiction loss when non-optimal scanners are used. In this section we answer the following question: Suppose that, for practical reasons for example, one uses a non-optimal scanner, accompanied with the optimal predictor for that scan. How large is the excess loss incurred by this scheme with respect to optimal scandiction?

For the sake of simplicity, we consider the scenario of predicting the next outcome of a binary source, with $D = [0, 1]$

as the prediction space. Hence, $l : \{0,1\} \times [0,1] \to \mathbb{R}$ is the loss function. Let ϕ_l denote the Bayes envelope associated with l, i.e.,

$$\phi_l(p) = \min_{q \in [0,1]} [(1-p)l(0,q) + pl(1,q)]. \quad (9)$$

We further define

$$\epsilon_l = \min_{\alpha,\beta} \max_{0 \le p \le 1} |\alpha h_b(p) + \beta - \phi_l(p)|, \quad (10)$$

where $h_b(p)$ is the binary entropy function. Thus ϵ_l is the error in approximating $\phi_l(p)$ by the best affine function of $h_b(p)$. For example, when l is the Hamming loss function, denoted by l_H, we have $\epsilon_{l_H} = 0.08$ and when l is the squared error, denoted by l_s, $\epsilon_{l_s} = 0.0137$. For the log loss, however, the expected instantaneous loss equals the conditional entropy, hence the expected cumulative loss coincides with the entropy, which is invariant to the scan, and we have $\epsilon_l = 0$. To wit, the scan is inconsequential under log loss.

Let Ψ be any (possibly data dependent) scan, and let $E_{Q_B} \frac{1}{|B|} L_{(\Psi, F^{opt})}(X_B)$ denote the expected normalized cumulative loss in scandicting X_B with the scan Ψ and the optimal predictor for that scan, under the loss function l. Remembering that $U(l, Q_B)$ denotes the scandictability of X_B w.r.t the loss function l, namely, $U(l, Q_B) = \inf_\Psi E_{Q_B} \frac{1}{|B|} L_{(\Psi, F^{opt})}(X_B)$, we have the following.

Theorem 5: Let X_B be an arbitrarily distributed binary field. Then, for any scan Ψ,

$$\left| E_{Q_B} \frac{1}{|B|} L_{(\Psi, F^{opt})}(X_B) - U(l, Q_B) \right| \le 2\epsilon_l. \quad (11)$$

That is, the excess loss incurred by applying *any scanner* Ψ, accompanied with the optimal predictor for that scan, with respect to optimal scandiction is not larger than $2\epsilon_l$.

Note that although the definitions of $\phi_l(p)$ and ϵ_l refer to the binary scenario, Theorem 5 holds for larger alphabets, with ϵ_l defined as in (10), with the maximum ranging over the simplex of all distributions on the alphabet, and $h(p)$ (replacing $h_b(p)$) and $\phi_l(p)$ denoting the entropy and Bayes envelope of the distribution p, respectively.

V. Prediction and Filtering of Noisy Data Arrays

In this section, we consider the scenario in which the predictor has access only to noisy observations of the data. We assume the joint probability distribution Q is known, and examine the setting of binary fields under Hamming loss (analogous results for Gaussian fields under squared error loss can be found in [11]). We characterize the noisy scandictability and the achieving scandictors in terms on the "clean" scandictability of the noisy data and give a bound on the excess loss when non optimal scanners are used (yet, with the optimal predictor for each scan). We then consider the interesting scenario of filtering, where the predictor is replaced by a filter, which has access to the current (noisy) observation as well. In that case, a lower bound on the best achievable performance is given. For the case of a binary valued field observed through a binary symmetric channel, we bound the excess loss when a non-optimal scanner is used (with an optimal filter) and show that in this scantering problem the excess loss is negligible and, in fact, the filtering performance is less sensitive to the scanning order (compared to the prediction counterpart).

We first formally define the noisy scenario. Let $\{(X_t, Y_t)\}_{t \in \mathbb{Z}^2}$ be a random field with components, (X_t, Y_t), in $A \times N$, where N is the noisy observation alphabet. Here, $\{X_t\}_{t \in \mathbb{Z}^2}$ represents the clean signal and $\{Y_t\}_{t \in \mathbb{Z}^2}$ represents the noisy observations. We assume that the noisy observations are stochastically connected to the clean signal.

The notion of scandiction is similar to that of Section II, however, both the scanner and the predictor are allowed to access only the noisy signal $\{Y_t\}$. Namely, for any $B \in \mathcal{V}$, the scan, Ψ, is a sequence of measurable mappings $\{\Psi_t\}_{t=1}^{|B|}, \Psi_t : A^{t-1} \to B$, determining the next site in B according to the previously observed values $\{Y_i\}_{i \in \Psi_1, \ldots, \Psi_{t-1}}$. The predictor, F, is a sequence of measurable mappings $\{F_t\}_{t=1}^{|B|}, F_t : A^{t-1} \to D$, determining the prediction for the value of x_{Ψ_t} according to the previously observed values $\{Y_i\}_{i \in \Psi_1, \ldots, \Psi_{t-1}}$. The cumulative loss of a scandictor (Ψ, F) is given by $L_{(\Psi,F)}(x_B, y_B)$, the sum of the instantaneous losses over the array B, i.e.,

$$L_{(\Psi,F)}(x_B, y_B) = \sum_{t=1}^{|B|} l(x_{\Psi_t}, F_t(y_{\Psi_1}, \ldots, y_{\Psi_{t-1}})). \quad (12)$$

For a given loss function l and source $Q \in \mathcal{M}(\Omega)$ restricted to B, define the *noisy scandictability* by

$$\bar{U}(l, Q_B) = \inf_{(\Psi,F) \in \mathcal{S}(B)} E_{Q_B} \frac{1}{|B|} L_{(\Psi,F)}(X_B, Y_B), \quad (13)$$

and the noisy scandictability for the source Q, $\bar{U}(l, Q)$, by

$$\bar{U}(l, Q) = \lim_{n \to \infty} \bar{U}(l, Q_{V_n}), \quad (14)$$

if this limit exists.

In the important case where F_t is allowed to base its estimation on y_{Ψ_t} as well, we denote it by $\tilde{F}_t$, and we have

$$L_{(\Psi,\tilde{F})}(x_B, y_B) = \sum_{t=1}^{|B|} l(x_{\Psi_t}, \tilde{F}_t(y_{\Psi_1}, \ldots, y_{\Psi_t})). \quad (15)$$

As scanning of the data is also involved in this filtering setting, we refer to this problem as *scantering*. We thus define

$$\tilde{U}(l, Q_B) = \inf_{(\Psi,\tilde{F})} E_{Q_B} \frac{1}{|B|} L_{(\Psi,\tilde{F})}(X_B, Y_B), \quad (16)$$

and

$$\tilde{U}(l, Q) = \lim_{n \to \infty} \tilde{U}(l, Q_{V_n}), \quad (17)$$

if this limit exists.

Analogously to [6, Theorem 1], it can be shown that for any stationary random field both the limit in (14) and that in (17) exist.

A. Scandiction of Noisy Data Arrays

We consider the binary scandiction problem, i.e., where $\{X_t\}$ is a binary random field passed through a binary symmetric channel with cross over probability $\delta < \frac{1}{2}$, and $\{Y_t\}$ is the channel output. The following proposition relates the noisy scandictability, $\bar{U}(l_H, Q)$, to the clean one, $U(l_H, Q_Y)$, where Q_Y is the marginal probability measure of Y, and gives a bound on the excess loss incurred when a non-optimal scanner is used.

Proposition 6: Let $\{(X_t, Y_t)\}_{t\in\mathbb{Z}^2}$ be a binary random field governed by a probability measure Q such that $\{Y_t\}$ is the output of a binary memoryless symmetric channel with cross over probability $\delta < 1/2$ and input $\{X_t\}$. Then, $\bar{U}(l_H, Q) = \frac{U(l_H,Q_Y)-\delta}{1-2\delta}$, and $\bar{U}(l_H, Q)$ is achieved by the scandictor which achieves $U(l_H, Q_Y)$. Furthermore, for any scandictor (Ψ, F^{opt}), where Ψ is arbitrary and F^{opt} is the optimal predictor for Ψ, we have

$$\left|\frac{1}{|B|}E_{Q_B}L_{\Psi,F^{opt}}(X_B, Y_B) - \bar{U}(l_H, Q_B)\right| \le \frac{2\epsilon_{l_H}}{1-2\delta}. \quad (18)$$

B. Scantering of Noisy Data Arrays

In this scenario, the filter, $\tilde{F}$, has access, in addition to $Y_{\Psi_1}, \ldots, Y_{\Psi_{t-1}}$, to Y_{Ψ_t} when estimating X_{Ψ_t}. We assume an invertible memoryless channel, meaning the input distribution of a single symbol is uniquely determined given the output distribution. In this case, the associated Bayes envelope is $\phi_l(P) = \min_{f(\cdot)} El(X, f(Y))$, where P is the distribution of the channel output Y. Define $\zeta(d) = \max\{H(P) : \phi_l(P) \le d\}$ and let $\bar{\zeta}(\cdot)$ be the upper concave envelope of $\zeta(\cdot)$. The following theorem is the direct analogue of the lower bounds in [6] for the filtering scenario. Note, however, that it holds for any finite n.

Theorem 7: Let Y_B be the output of an invertible memoryless channel whose input is X_B. Then, for any scanterer $(\Psi, \tilde{F})$ we have

$$\bar{\zeta}\left(\frac{1}{|B|}E_{Q_B}L_{\Psi,\tilde{F}}(X_B, Y_B)\right) \ge \frac{1}{|B|}H(Y_B). \quad (19)$$

In particular, $\bar{\zeta}\left(\tilde{U}(l, Q_B)\right) \ge \frac{1}{|B|}H(Y_B)$, where $H(Y_B)$ is the entropy of Y_B.

Theorem 7 lower bounds the best possible scantering performance. Similarly to Section IV, it is interesting to derive a bound on the excess loss when a non-optimal scanning order is used (yet with the optimal filter for that scan). Define

$$f_\delta(p) = \min\left\{\frac{p-\delta}{1-2\delta}, \delta\right\} \quad (20)$$

and

$$\epsilon_\delta = \min_{a,b} \max_{\delta\le p\le 1/2} \left|ah_b(p) + b - f_\delta(p)\right|. \quad (21)$$

Under these definitions, we have the following theorem.

Theorem 8: Let Y_B be the output of a binary symmetric channel with crossover probability δ whose input is X_B. Then, for any scanterer $(\Psi, \tilde{F}^{opt})$, where $\tilde{F}^{opt}$ is the optimal filter for the scan Ψ, we have

$$\left|\frac{1}{|B|}E_{Q_B}L_{\Psi,\tilde{F}^{opt}}(X_B, Y_B) - \tilde{U}(l_H, Q_B)\right| \le 2\epsilon_\delta. \quad (22)$$

Even without evaluating ϵ_δ explicitly, it is easy to see that the excess loss when using non optimal scanners is quite small in the filtering scenario. For example, for $\delta = 0.1$ and $\delta = 0.25$ we have $\epsilon_\delta < 0.035$ and $\epsilon_\delta < 0.03$ respectively, yielding a maximal loss of 0.07 or even 0.06. This should be compared to 0.16 in the prediction scenario (or even larger values in the noisy prediction scenario). The fact that the filtering problem is less sensitive to the scanning order is quite clear as the noisy observation of X_{Ψ_t} is available under any scan.

VI. Conclusion

In this paper, we formally defined finite set scandictability, and showed that there exists a universal algorithm which successfully competes with any finite set of scandictors when the random field is stationary. Moreover, the existence of a universal algorithm which achieves the scandictability of any spatially stationary random field was also established.

We then considered the scenario where non-optimal scanners are used, and derived a bound on the excess loss in that case, compared to optimal scandiction.

Finally, the noisy scenario was discussed, in which the scandictor (or scanterer) has access only to a noisy observation of the data. For the interesting case of scantering, a bound on the best scantering performance was given, as well as a bound on the excess loss incurred when a non-optimal scanner is used.

References

[1] M.J. Weinberger, G. Seroussi, and G. Sapiro, "LOCO-I: A low complexity, context-based, lossless image compression algorithm," *Proc. IEEE Data Compression Conf.*, pp. 140–149, 1996.

[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] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inform. Theory*, vol. IT-24, pp. 530–536, September 1978.

[4] T. Weissman and S. Mannor, "On universal compression of multi-dimensional data arrays using self-similar curves," in *Proc. 38th Annu. Allerton Conf. Communication, Control, and Computing*, October 2000, vol. I, pp. 470–479.

[5] 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.

[6] 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.

[7] 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.

[8] V. G. Vovk, "Aggregating strategies," *Proc. 3rd Annu. Workshop Computational Learning Theory,* San Mateo, CA, pp. 372–383, 1990.

[9] N. Merhav, E. Ordentlich, G. Seroussi, and M. J. Weinberger, "On sequential strategies for loss functions with memory," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1947–1958, July 2002.

[10] T. Weissman and N. Merhav, "On limited-delay lossy coding and filtering of individual sequences," *IEEE Trans. Inform. Theory*, vol. 48, no. 3, pp. 721–733, March 2002.

[11] A. Cohen, T. Weissman, and N. Merhav, "Universal scanning and sequential decision making for multi-dimensional data," *In preperation.*