Scanning and sequential decision making for multidimensional data

Asaf Cohen, Neri Merhav and Tsachy Weissman

CMI Retreat, La Canada, October 2007

Introduction

Definition (Informal)

Sequentially scan a multidimensional data array, while minimizing a given cumulative loss function.

Motivation

- Image and video processing: Predictive coding or filtering.
- Halftoning, 1-D wavelet processing, or pattern recognition/matching.
- Space filling curves for database query, image rendering and more.

Scandiction Defined

Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

- $l : A \times A \longrightarrow [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.

Scandiction Defined

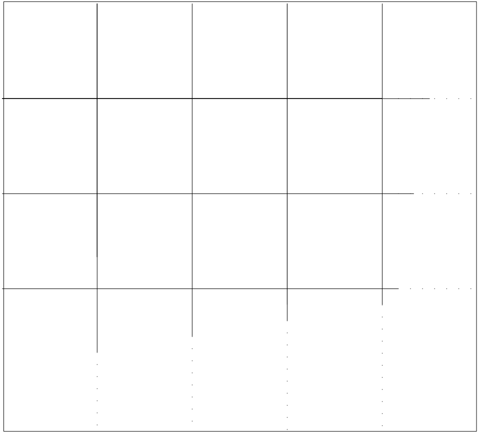
Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

- $l : A \times A \longrightarrow [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.



Scandiction Defined

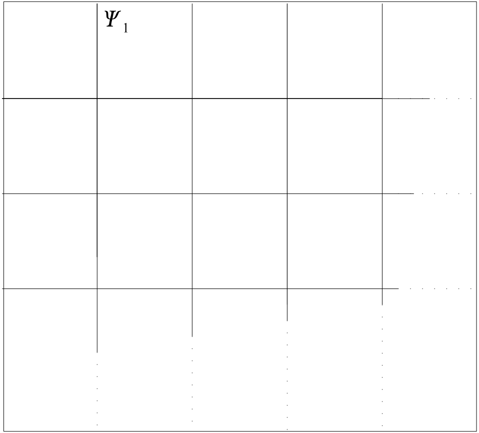
Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

- $l : A \times A \longrightarrow [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.



Scandiction Defined

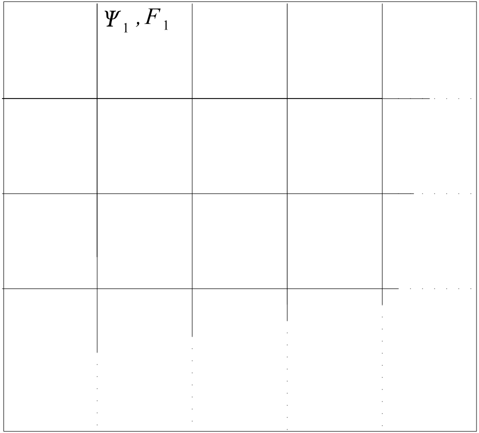
Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

- $l : A \times A \longrightarrow [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.



Scandiction Defined

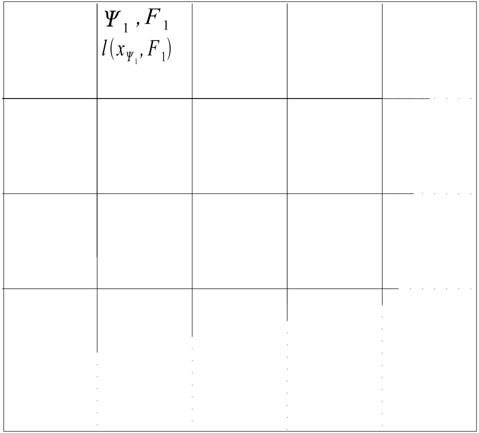
Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

- $l : A \times A \longrightarrow [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.



Scandiction Defined

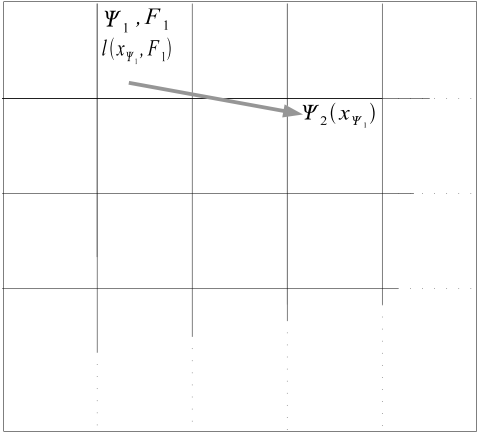
Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

- $l : A \times A \longrightarrow [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.



Scandiction Defined

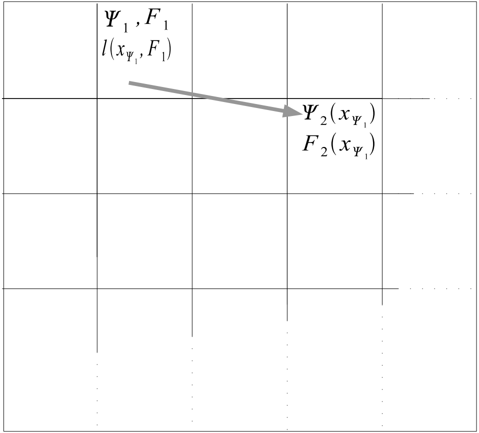
Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

- $l : A \times A \longrightarrow [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.



Scandiction Defined

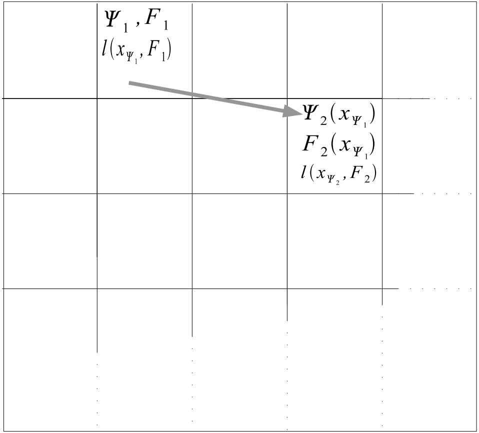
Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

- $l : A \times A \rightarrow [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.



Scandiction Defined

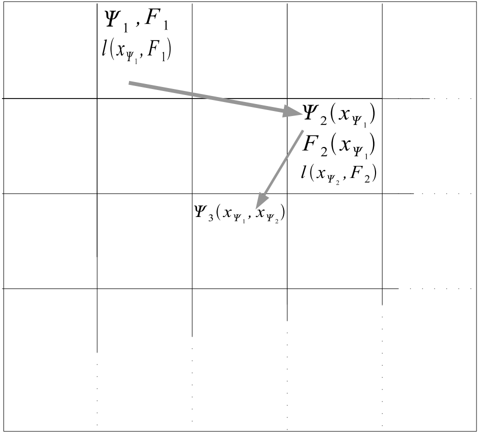
Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

- $l : A \times A \longrightarrow [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.



Scandiction Defined

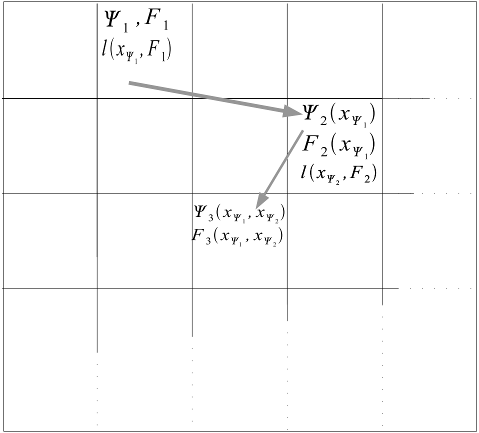
Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

- $l : A \times A \longrightarrow [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.



Scandiction Defined

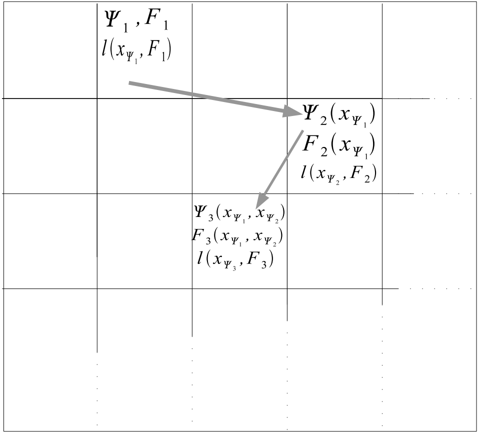
Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

- $l : A \times A \to [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.



Scandiction Defined

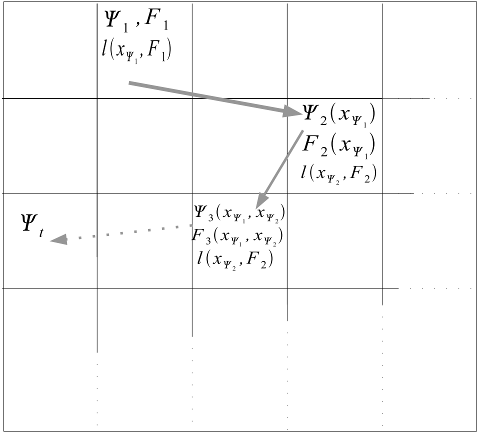
Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

- $l : A \times A \longrightarrow [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.



Scandiction Defined

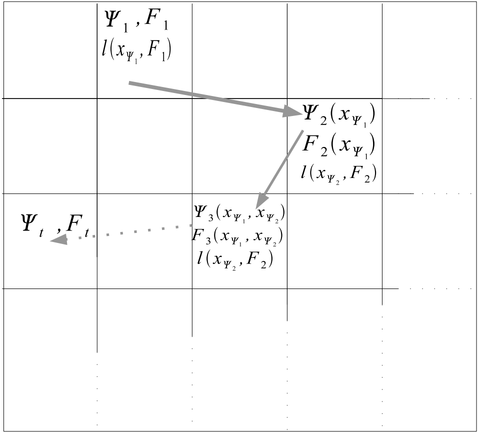
Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

- $l : A \times A \longrightarrow [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.



Scandiction Defined

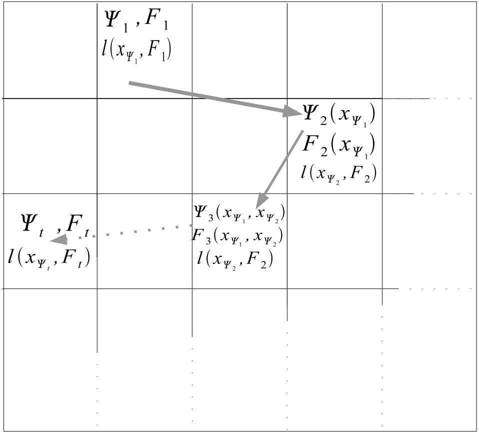
Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

- $l : A \times A \longrightarrow [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.



Scandiction Defined

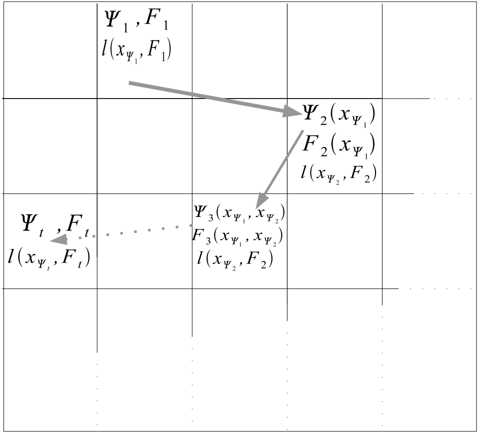
Definition

A *Scandictor* (Ψ, F):

- $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$, $\Psi_t(x_{\Psi_1}^{\Psi_{t-1}})$ determining the site to be visited at time t.
- $\{F_t\}_{t=1}^{|B|}$, $F_t : A^{t-1} \mapsto D$, $F_t(x_{\Psi_1}^{\Psi_{t-1}})$ is the prediction for the value at Ψ_t.
- Randomized scandictors.

The loss function:

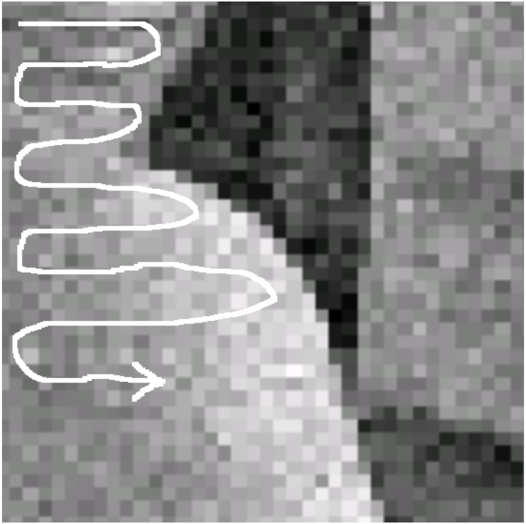
- $l : A \times A \rightarrow [0, \infty)$ is a bounded loss function.
- $L_{(\Psi,F)}(x_{V_n})$ is the cumulative loss of (Ψ, F) over x_{V_n}.



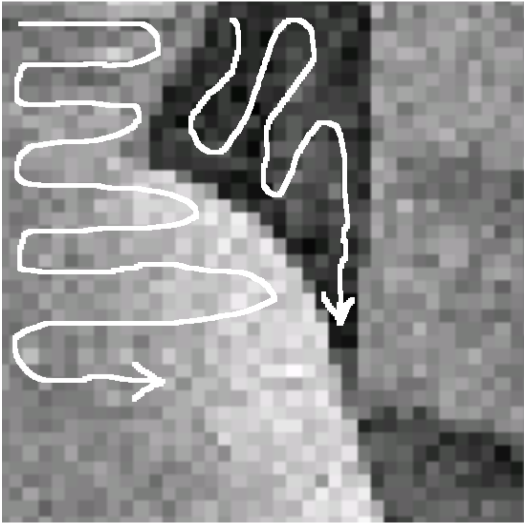
Scanning of a Noisy Image

Scanning of a Noisy Image

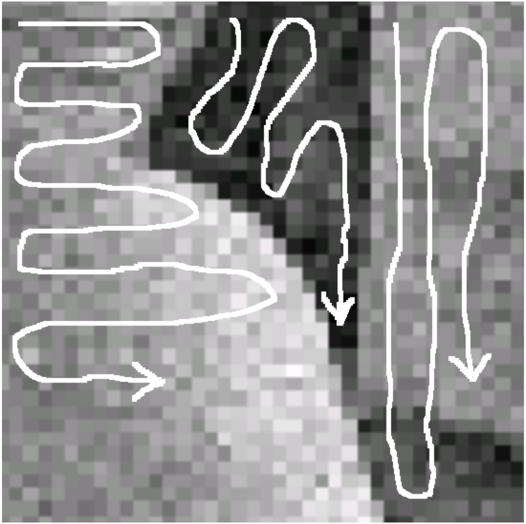
Scanning of a Noisy Image



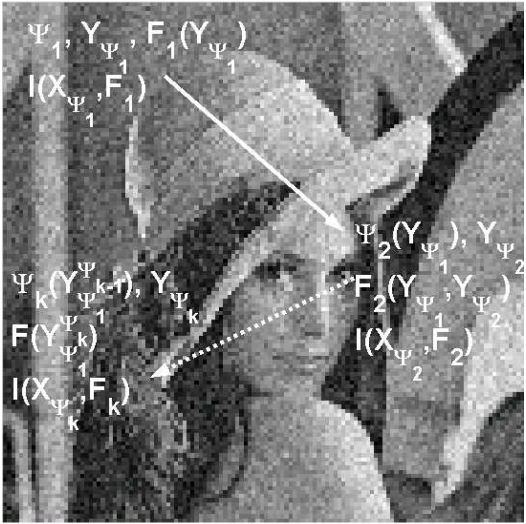
Scanning of a Noisy Image



Scanning of a Noisy Image



Scanning of a Noisy Image



- $\{(X_t, Y_t)\}_{t \in B}$, $B \subset \mathbb{Z}^2$, is a random field
- Scanner: $\{\Psi_t\}_{t=1}^{|B|}$, $\Psi_t : A^{t-1} \mapsto B$
- Filter: $\{F_t\}_{t=1}^{|B|}$, $F_t : A^t \mapsto D$
- Predictor: $F_t : A^{t-1} \mapsto D$
- Loss: $l : A \times D \rightarrow [0, \infty)$
- Cumulative loss: $L_{(\Psi,F)}(x_B, y_B)$

Unanswered Questions

1. Given a random field with a known distribution Q, what is the optimal scan?
2. Given a random field with an unknown distribution, can we construct a universal scan?
3. Can one compete successfully with a finite set of scanners on any individual image?
   - Any two arbitrary scanners.
   - Any two finite state scanners.
   - Some finite set of scanners.
4. What is the loss in using a non-optimal scanner?
5. What if the observations are corrupted in some way?
6. What if resources (e.g., memory or flops) are limited?

Unanswered Questions

1. Given a random field with a known distribution Q, what is the optimal scan?
2. Given a random field with an unknown distribution, can we construct a universal scan?
3. Can one compete successfully with a finite set of scanners on any individual image?
    - Any two arbitrary scanners.
    - Any two finite state scanners.
    - Some finite set of scanners.
4. What is the loss in using a non-optimal scanner?
5. What if the observations are corrupted in some way?
6. What if resources (e.g., memory or flops) are limited?

Unanswered Questions

1. Given a random field with a known distribution Q, what is the optimal scan?
2. Given a random field with an unknown distribution, can we construct a universal scan?
3. Can one compete successfully with a finite set of scanners on any individual image?
   - Any two arbitrary scanners.
   - Any two finite state scanners.
   - Some finite set of scanners.
4. What is the loss in using a non-optimal scanner?
5. What if the observations are corrupted in some way?
6. What if resources (e.g., memory or flops) are limited?

Unanswered Questions

1. Given a random field with a known distribution Q, what is the optimal scan?
2. Given a random field with an unknown distribution, can we construct a universal scan?
3. Can one compete successfully with a finite set of scanners on any individual image?
    - Any two arbitrary scanners.
    - Any two finite state scanners.
    - Some finite set of scanners.
4. What is the loss in using a non-optimal scanner?
5. What if the observations are corrupted in some way?
6. What if resources (e.g., memory or flops) are limited?

Unanswered Questions

1. Given a random field with a known distribution Q, what is the optimal scan?
2. Given a random field with an unknown distribution, can we construct a universal scan?
3. Can one compete successfully with a finite set of scanners on any individual image?
   - Any two arbitrary scanners.
   - Any two finite state scanners.
   - Some finite set of scanners.
4. What is the loss in using a non-optimal scanner?
5. What if the observations are corrupted in some way?
6. What if resources (e.g., memory or flops) are limited?

Unanswered Questions

1. Given a random field with a known distribution Q, what is the optimal scan?
2. Given a random field with an unknown distribution, can we construct a universal scan?
3. Can one compete successfully with a finite set of scanners on any individual image?
    - Any two arbitrary scanners.
    - Any two finite state scanners.
    - Some finite set of scanners.
4. What is the loss in using a non-optimal scanner?
5. What if the observations are corrupted in some way?
6. What if resources (e.g., memory or flops) are limited?

Unanswered Questions

1. Given a random field with a known distribution Q, what is the optimal scan?
2. Given a random field with an unknown distribution, can we construct a universal scan?
3. Can one compete successfully with a finite set of scanners on any individual image?
   - Any two arbitrary scanners.
   - Any two finite state scanners.
   - Some finite set of scanners.
4. What is the loss in using a non-optimal scanner?
5. What if the observations are corrupted in some way?
6. What if resources (e.g., memory or flops) are limited?

Unanswered Questions

1. Given a random field with a known distribution Q, what is the optimal scan?
2. Given a random field with an unknown distribution, can we construct a universal scan?
3. Can one compete successfully with a finite set of scanners on any individual image?
   - Any two arbitrary scanners.
   - Any two finite state scanners.
   - Some finite set of scanners.
4. What is the loss in using a non-optimal scanner?
5. What if the observations are corrupted in some way?
6. What if resources (e.g., memory or flops) are limited?

Outline

Universal Scanning Compared to Universal Prediction

Example (Real number followed by its binary representation)

$Y_1 \sim U[0,1], \; Y_2^n \in \{0,1\}^{n-1}.$

Y_1	0	1	1	0	1	. . . .

Claims

Under the square error loss:

- For any scandictor (Ψ, F), $EL_{(\Psi,F)}(X^n) \geq \frac{n-1}{8}$.
- $E \min\{L_{(\text{trivial},F)}(X^n), L_{(\text{reversed},F)}(X^n)\} \leq \frac{n}{16} + o(n)$.

Theorem

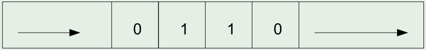
There exist two scandictors $(\Psi, F)_1$ and $(\Psi, F)_2$, such that for any scandictor (Ψ, F) there exists x^n for which

$$L_{(\Psi,F)}(x^n) - \min\{L_{(\Psi,F)_1}(x^n), L_{(\Psi,F)_2}(x^n)\} = \Theta(n).$$

Universal Scanning Compared to Universal Prediction

Example (Real number followed by its binary representation)

Random cyclic shift right.



Claims

Under the square error loss:

- For any scandictor (Ψ, F), $EL_{(\Psi,F)}(X^n) \geq \frac{n-1}{8}$.
- $E \min\{L_{(\text{trivial},F)}(X^n), L_{(\text{reversed},F)}(X^n)\} \leq \frac{n}{16} + o(n)$.

Theorem

There exist two scandictors $(\Psi, F)_1$ and $(\Psi, F)_2$, such that for any scandictor (Ψ, F) there exists x^n for which

$$L_{(\Psi,F)}(x^n) - \min\{L_{(\Psi,F)_1}(x^n), L_{(\Psi,F)_2}(x^n)\} = \Theta(n).$$

Universal Scanning Compared to Universal Prediction

Example (Real number followed by its binary representation)

$X_J = Y_1, \ J \sim U\{0, 1, \ldots, n-1\}$

......	Y_1	0	1	1	0	1	....

Claims

Under the square error loss:

- For any scandictor (Ψ, F), $EL_{(\Psi,F)}(X^n) \geq \frac{n-1}{8}$.
- $E \min\{L_{(\text{trivial},F)}(X^n), L_{(\text{reversed},F)}(X^n)\} \leq \frac{n}{16} + o(n)$.

Theorem

There exist two scandictors $(\Psi, F)_1$ and $(\Psi, F)_2$, such that for any scandictor (Ψ, F) there exists x^n for which

$$L_{(\Psi,F)}(x^n) - \min\{L_{(\Psi,F)_1}(x^n), L_{(\Psi,F)_2}(x^n)\} = \Theta(n).$$

Universal Scanning Compared to Universal Prediction

Example (Real number followed by its binary representation)

$X_J = Y_1,\ J \sim U\{0, 1, \ldots, n-1\}$

. . . . .	Y_1	0	1	1	0	1	. . . . . .

Claims

Under the square error loss:

- For any scandictor (Ψ, F), $EL_{(\Psi,F)}(X^n) \geq \frac{n-1}{8}$.
- $E \min\{L_{(\text{trivial},F)}(X^n), L_{(\text{reversed},F)}(X^n)\} \leq \frac{n}{16} + o(n)$.

Theorem

There exist two scandictors $(\Psi, F)_1$ and $(\Psi, F)_2$, such that for any scandictor (Ψ, F) there exists x^n for which

$$L_{(\Psi,F)}(x^n) - \min\{L_{(\Psi,F)_1}(x^n), L_{(\Psi,F)_2}(x^n)\} = \Theta(n).$$

Universal Scanning Compared to Universal Prediction

Example (Real number followed by its binary representation)

$X_J = Y_1,\ J \sim U\{0, 1, \ldots, n-1\}$

. . . . .	Y_1	0	1	1	0	1	. . . . . .

Claims

Under the square error loss:

- For any scandictor (Ψ, F), $EL_{(\Psi,F)}(X^n) \geq \frac{n-1}{8}$.
- $E \min\{L_{(\text{trivial},F)}(X^n), L_{(\text{reversed},F)}(X^n)\} \leq \frac{n}{16} + o(n)$.

Theorem

There exist two scandictors $(\Psi, F)_1$ and $(\Psi, F)_2$, such that for any scandictor (Ψ, F) there exists x^n for which

$$L_{(\Psi,F)}(x^n) - \min\{L_{(\Psi,F)_1}(x^n), L_{(\Psi,F)_2}(x^n)\} = \Theta(n).$$

Scandictability

Definition (Merhav and Weissman 03)

The scandictability of a stationary source governed by Q, $U(l, Q)$, is the best achievable performance for Q. More specifically,

$$U(l, Q) = \lim_{n\to\infty} \inf_{(\Psi,F)\in\mathcal{S}(V_n)} E_Q \frac{1}{n^2} L_{(\Psi,F)}(X_{V_n}),$$

where $\mathcal{S}(V_n)$ is the set of *all* possible scandictors for V_n.

In the stochastic setting, we expect to achieve $U(l, Q)$ (in some sense) under mild conditions on the random field.

Universal Scandiction

Theorem

Let X be a stationary random field with a probability measure Q. Let $\mathcal{F}$ be a finite set of scandictors. Then, there exists a scandictor $(\hat{\Psi}, \hat{F})$, independent of Q, for which

$$\liminf_{n\to\infty} EL_{(\hat{\Psi},\hat{F})}(X_{V_n}) \leq \liminf_{n\to\infty} \min_{(\Psi,F)\in\mathcal{F}} E_Q L_{(\Psi,F)}(X_{V_n}).$$

That is, there exists a universal scandictor which successfully competes with any finite set of scandictors.

Proof outline.

Force all scandictors to pass through "check-points" in the image.
Probe their performance at the check-points.

- For any "Individual image":
  Check-points are dense enough -
  it is possible to successfully compete with any finite set.
- Any spatially stationary field:
  Check points are not too dense -
  each scandictor's performance is essentially the same as the unlimited version.

Proof outline.

Force all scandictors to pass through "check-points" in the image.
Probe their performance at the check-points.

- For any "Individual image":
  Check-points are dense enough -
  it is possible to successfully compete with any finite set.
- Any spatially stationary field:
  Check points are not too dense -
  each scandictor's performance is essentially the same as the unlimited version.

Proof outline.

Force all scandictors to pass through "check-points" in the image.
Probe their performance at the check-points.

- For any "Individual image":
  Check-points are dense enough -
  it is possible to successfully compete with any finite set.
- Any spatially stationary field:
  Check points are not too dense -
  each scandictor's performance is essentially the same as the unlimited version.

□

Is $U(l, Q)$ is Achievable?

Question

Can we find a set of scandictors, which is not too large to compete with, yet is rich enough to asymptotically cover interesting random fields?

Definition (Lempel and Ziv 86)

A finite state scandictor is a scandictor for which: $\Psi_{t+1} = \Psi_t + d(s_t)$, $F_{t+1} = F(s_t)$ and $s_t = g(s_{t-1}, x_{\Psi_t})$.

Lemma

Let $\mathcal{F}_S = \{(\Psi, F)_j\}$ be the set of all finite-state scandictors with at most S states. Then, for any stationary random field Q,

$$\lim_{S\to\infty} U_{\mathcal{F}_S}(l, Q) = U(l, Q).$$

Is $U(l, Q)$ is Achievable?

Question

Can we find a set of scandictors, which is not too large to compete with, yet is rich enough to asymptotically cover interesting random fields?

Definition (Lempel and Ziv 86)

A finite state scandictor is a scandictor for which: $\Psi_{t+1} = \Psi_t + d(s_t)$, $F_{t+1} = F(s_t)$ and $s_t = g(s_{t-1}, x_{\Psi_t})$.

Lemma

Let $\mathcal{F}_S = \{(\Psi, F)_j\}$ be the set of all finite-state scandictors with at most S states. Then, for any stationary random field Q,

$$\lim_{S \to \infty} U_{\mathcal{F}_S}(l, Q) = U(l, Q).$$

Is $U(I, Q)$ is Achievable?

Question

Can we find a set of scandictors, which is not too large to compete with, yet is rich enough to asymptotically cover interesting random fields?

Definition (Lempel and Ziv 86)

A finite state scandictor is a scandictor for which: $\Psi_{t+1} = \Psi_t + d(s_t)$, $F_{t+1} = F(s_t)$ and $s_t = g(s_{t-1}, x_{\Psi_t})$.

Lemma

Let $\mathcal{F}_S = \{(\Psi, F)_j\}$ be the set of all finite-state scandictors with at most S states. Then, for any stationary random field Q,

$$\lim_{S\to\infty} U_{\mathcal{F}_S}(I, Q) = U(I, Q).$$

Achievability of $U(I, Q)$

Theorem

Let X be a stationary random field over a finite alphabet A and a probability measure Q. Then, there exists a scandictor (Ψ, F), independent of Q, for which

$$\lim_{n \to \infty} EL_{(\Psi,F)}(X_{V_n}) = U(I, Q).$$

Outline

The Excess Loss When Non-Optimal Scanners are Used

The existence of a universal scandictor which achieves $U(I, Q)$ was established. However, it is interesting to ask, from both practical and theoretical reasons, the following question:

Question

What is the excess scandiction loss when non-optimal scanners are used?

Excess Loss - Results

Theorem

Let X_B be an arbitrarily distributed binary field. Then, for any scandictor (Ψ, F^{opt}), where F^{opt} is the optimal predictor for Ψ,

$$\left| E_Q L_{(\Psi, F^{opt})}(X_B) - U(l, Q_B) \right| \le 2\epsilon_l.$$

Examples

For Hamming loss, $\epsilon_{l_H} = 0.08$. For squared error, $\epsilon_{l_s} = 0.0137$, and for log loss, $\epsilon_{l_{log}} = 0$.

Tightness

There *are* scans which perform badly!

Excess Loss - Results

Theorem

Let X_B be an arbitrarily distributed binary field. Then, for any scandictor (Ψ, F^{opt}), where F^{opt} is the optimal predictor for Ψ,

$$\left| E_Q L_{(\Psi, F^{opt})}(X_B) - U(l, Q_B) \right| \leq 2\epsilon_l.$$

Examples

For Hamming loss, $\epsilon_{l_H} = 0.08$. For squared error, $\epsilon_{l_s} = 0.0137$, and for log loss, $\epsilon_{l_{log}} = 0$.

Tightness

There *are* scans which perform badly!

Universality in the Individual Setting

Question

There are strong universality results in the stochastic setting. What about the individual setting?

- No stationarity or block ergodicity: performance of block-wise scandictors does not converge to the optimum.
- Cannot compete with any two scandictors.
- No immediate feature which is scan-invariant.

What can we expect?

- Universality results for limited sets of scandictors (e.g., raster-type scans with Markov predictors).
- Find scans with uniformly small (but not vanishing) redundancy with respect to some interesting set of scans.

Universality in the Individual Setting

Question

There are strong universality results in the stochastic setting. What about the individual setting?

- No stationarity or block ergodicity: performance of block-wise scandictors does not converge to the optimum.
- Cannot compete with any two scandictors.
- No immediate feature which is scan-invariant.

What can we expect?

- Universality results for limited sets of scandictors (e.g., raster-type scans with Markov predictors).
- Find scans with uniformly small (but not vanishing) redundancy with respect to some interesting set of scans.

The Performance of the Peano-Hilbert Scan

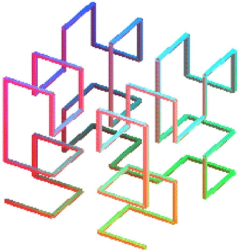
Theorem

Let x be any individual image. Let PH denote the Peano-Hilbert sequence of scans. Then, for any sequence of finite state scans Ψ and any loss function $l : \{0,1\} \times [0,1] \rightarrow \mathbb{R}$,

$$L_{PH}(x) \leq L_{\Psi}(x) + 2\epsilon_l .$$

Corollary (Hamming Loss)

$$Ł^{H}_{PH}(x) \leq L^{H}_{\Psi}(x) + \frac{1}{2}\rho(x) - h^{-1}(\rho(x)).$$



Outline

Definitions

$\{(X_t, Y_t)\}_{t\in\mathbb{Z}^2}$ is a random field, $(X_t, Y_t) \in A \times N$, 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.

Motivation

- Image and video filtering.
- Compression of noisy data.

Definition (Noisy Scandictability and Scantering)

- Y_t is available (filtering)- "scantering", $\tilde{U}(I, Q)$.
- Y_t is not available (prediction)- "noisy scandictability", $\bar{U}(I, Q)$.

A Bound on the Best Achievable Scanning and Filtering Performance

For an invertible memoryless channel, define

$$\phi_l(P) = \min_{f(\cdot)} El(X, f(Y)), \quad Y \sim P$$
$$\zeta(d) = \max\{H(P) : \phi_l(P) \leq d\}.$$

Let $\bar{\zeta}(\cdot)$ be the upper concave envelope of $\zeta(\cdot)$. Note: $\bar{\zeta}$ is a single letter expression.

Theorem

Assume an invertible memoryless channel. Then, for any (Ψ, F),

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

Gaussian Channel

- Simple expression for Gaussian input, but still not tight.
- We currently cannot identify the optimal scan.

A Bound on the Best Achievable Scanning and Filtering Performance

For an invertible memoryless channel, define

$$\begin{aligned}\phi_l(P) &= \min_{f(\cdot)} El(X, f(Y)), \quad Y \sim P \\ \zeta(d) &= \max\{H(P) : \phi_l(P) \leq d\}.\end{aligned}$$

Let $\bar{\zeta}(\cdot)$ be the upper concave envelope of $\zeta(\cdot)$. Note: $\bar{\zeta}$ is a single letter expression.

Theorem

Assume an invertible memoryless channel. Then, for any (Ψ, F),

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

Gaussian Channel

- Simple expression for Gaussian input, but still not tight.
- We currently cannot identify the optimal scan.

Excess Loss Bounds - Gaussian Setting

Theorem

Let X_B be Gaussian, with $Var(X_i) = \sigma_X^2 < \infty$. Let $Y_i = X_i + N_i$, where N_B is a WGN of variance σ_N^2. Then, for any Ψ^1 and Ψ^2,

$$\frac{1}{|B|}\left|EL_{(\Psi^1,F^{opt})}(X_B,Y_B) - EL_{(\Psi^2,F^{opt})}(X_B,Y_B)\right| \leq \sigma_X^2 \frac{f(SNR)}{SNR},$$

where $f(x) = \ln(1+x) - \frac{x}{x+1}$ and $SNR = \sigma_X^2/\sigma_N^2$.

- Single letter: $\sigma_X^2 \frac{f(SNR)}{SNR} = 2\sigma_N^2 I(SNR) - \sigma_X^2 mmse(SNR)$
- $\lim_{SNR\to 0^+} \frac{f(SNR)}{SNR} = \lim_{SNR\to\infty} \frac{f(SNR)}{SNR} = 0.$
- Slow decay at high SNR.

Excess Loss Bounds - Gaussian Setting

Theorem

Let X_B be Gaussian, with $Var(X_i) = \sigma_X^2 < \infty$. Let $Y_i = X_i + N_i$, where N_B is a WGN of variance σ_N^2. Then, for any Ψ^1 and Ψ^2,

$$\frac{1}{|B|}\left|EL_{(\Psi^1,F^{opt})}(X_B, Y_B) - EL_{(\Psi^2,F^{opt})}(X_B, Y_B)\right| \leq \sigma_X^2 \frac{f(SNR)}{SNR},$$

where $f(x) = \ln(1+x) - \frac{x}{x+1}$ and $SNR = \sigma_X^2/\sigma_N^2$.

- Single letter: $\sigma_X^2 \frac{f(\mathrm{SNR})}{\mathrm{SNR}} = 2\sigma_N^2 \mathrm{I}(\mathrm{SNR}) - \sigma_X^2 \mathrm{mmse}(\mathrm{SNR})$
- $\lim_{\mathrm{SNR}\to 0^+} \frac{f(\mathrm{SNR})}{\mathrm{SNR}} = \lim_{\mathrm{SNR}\to\infty} \frac{f(\mathrm{SNR})}{\mathrm{SNR}} = 0.$
- Slow decay at high SNR.

Excess Loss Bounds - Gaussian Setting

Theorem

Let X_B be Gaussian, with $Var(X_i) = \sigma_X^2 < \infty$. Let $Y_i = X_i + N_i$, where N_B is a WGN of variance σ_N^2. Then, for any Ψ^1 and Ψ^2,

$$\frac{1}{|B|}\left|EL_{(\Psi^1,F^{opt})}(X_B,Y_B) - EL_{(\Psi^2,F^{opt})}(X_B,Y_B)\right| \leq \sigma_X^2 \frac{f(SNR)}{SNR},$$

where $f(x) = \ln(1+x) - \frac{x}{x+1}$ and $SNR = \sigma_X^2/\sigma_N^2$.

- Single letter: $\sigma_X^2 \frac{f(\mathrm{SNR})}{\mathrm{SNR}} = 2\sigma_N^2 \mathrm{I}(\mathrm{SNR}) - \sigma_X^2 \mathrm{mmse}(\mathrm{SNR})$
- $\lim_{\mathrm{SNR}\to 0^+} \frac{f(\mathrm{SNR})}{\mathrm{SNR}} = \lim_{\mathrm{SNR}\to\infty} \frac{f(\mathrm{SNR})}{\mathrm{SNR}} = 0.$
- Slow decay at high SNR.

Excess Loss Bounds - Gaussian Setting

Theorem

Let X_B be Gaussian, with $Var(X_i) = \sigma_X^2 < \infty$. Let $Y_i = X_i + N_i$, where N_B is a WGN of variance σ_N^2. Then, for any Ψ^1 and Ψ^2,

$$\frac{1}{|B|}\left|EL_{(\Psi^1,F^{opt})}(X_B,Y_B) - EL_{(\Psi^2,F^{opt})}(X_B,Y_B)\right| \leq \sigma_X^2 \frac{f(SNR)}{SNR},$$

where $f(x) = \ln(1+x) - \frac{x}{x+1}$ and $SNR = \sigma_X^2/\sigma_N^2$.

- Single letter: $\sigma_X^2 \frac{f(\mathrm{SNR})}{\mathrm{SNR}} = 2\sigma_N^2 \mathrm{I}(\mathrm{SNR}) - \sigma_X^2 \mathrm{mmse}(\mathrm{SNR})$
- $\lim_{\mathrm{SNR}\to 0^+} \frac{f(\mathrm{SNR})}{\mathrm{SNR}} = \lim_{\mathrm{SNR}\to\infty} \frac{f(\mathrm{SNR})}{\mathrm{SNR}} = 0.$
- Slow decay at high SNR.

Excess Loss Bounds - Main Tools

Theorem (Duncan 70; Guo, Shamai and Verdu 05)

For any continuous input process of finite average power and AWGN,

$$I(SNR) = \frac{SNR}{2T} \int_0^T cmmse(SNR, t)dt.$$

- Right hand side: (continuous time) cmmse!
- Left hand side: (continuous time) mutual information.
- In discrete time: $I(X^n; Y^n)$ is scan invariant!

Analogous results can be derived for Binary input and AWGN:

$$\frac{1}{|B|} \left| EL_{(\Psi^1, F^{opt})}(X_B, Y_B) - EL_{(\Psi^2, F^{opt})}(X_B, Y_B) \right| \leq 2\sigma_N^2 \mathrm{I(SNR)} - \sigma_X^2 \mathrm{mmse(SNR)}.$$

Excess Loss Bounds - Main Tools

Theorem (Duncan 70; Guo, Shamai and Verdu 05)

For any continuous input process of finite average power and AWGN,

$$I(SNR) = \frac{SNR}{2T}\int_0^T cmmse(SNR,t)dt.$$

- Right hand side: (continuous time) cmmse!
- Left hand side: (continuous time) mutual information.
- In discrete time: $I(X^n;Y^n)$ is scan invariant!

Analogous results can be derived for Binary input and AWGN:

$$\frac{1}{|B|}\left|EL_{(\Psi^1,F^{opt})}(X_B,Y_B) - EL_{(\Psi^2,F^{opt})}(X_B,Y_B)\right| \leq 2\sigma_N^2 \mathrm{I}(\mathrm{SNR}) - \sigma_X^2 \mathrm{mmse}(\mathrm{SNR}).$$

Excess Loss Bounds - Main Tools

Theorem (Duncan 70; Guo, Shamai and Verdu 05)

For any continuous input process of finite average power and AWGN,

$$I(SNR) = \frac{SNR}{2T} \int_0^T cmmse(SNR, t)dt.$$

- Right hand side: (continuous time) cmmse!
- Left hand side: (continuous time) mutual information.
- In discrete time: $I(X^n; Y^n)$ is scan invariant!

Analogous results can be derived for Binary input and AWGN:

$$\frac{1}{|B|}\left|EL_{(\Psi^1, F^{opt})}(X_B, Y_B) - EL_{(\Psi^2, F^{opt})}(X_B, Y_B)\right| \leq 2\sigma_N^2 \mathrm{I(SNR)} - \sigma_X^2 \mathrm{mmse(SNR)}.$$

Excess Loss Bounds - Main Tools

Theorem (Duncan 70; Guo, Shamai and Verdu 05)

For any continuous input process of finite average power and AWGN,

$$I(SNR) = \frac{SNR}{2T}\int_0^T cmmse(SNR, t)dt.$$

- Right hand side: (continuous time) cmmse!
- Left hand side: (continuous time) mutual information.
- In discrete time: $I(X^n; Y^n)$ is scan invariant!

Analogous results can be derived for Binary input and AWGN:

$$\frac{1}{|B|}\left|EL_{(\Psi^1,F^{opt})}(X_B,Y_B) - EL_{(\Psi^2,F^{opt})}(X_B,Y_B)\right| \leq 2\sigma_N^2 \mathrm{I(SNR)} - \sigma_X^2 \mathrm{mmse(SNR)}.$$

Excess Loss Bounds - Main Tools

Theorem (Duncan 70; Guo, Shamai and Verdu 05)

For any continuous input process of finite average power and AWGN,

$$I(SNR) = \frac{SNR}{2T} \int_0^T cmmse(SNR, t) dt.$$

- Right hand side: (continuous time) cmmse!
- Left hand side: (continuous time) mutual information.
- In discrete time: $I(X^n; Y^n)$ is scan invariant!

Analogous results can be derived for Binary input and AWGN:

$$\frac{1}{|B|} \left| EL_{(\Psi^1, F^{opt})}(X_B, Y_B) - EL_{(\Psi^2, F^{opt})}(X_B, Y_B) \right| \leq 2\sigma_N^2 \mathrm{I}(\mathrm{SNR}) - \sigma_X^2 \mathrm{mmse}(\mathrm{SNR}).$$

Excess Loss Bounds - Binary Input and BSC

Theorem

Let Y_t be the output of a BSC(δ) whose input is X_t. Then, for any scanterer $(\Psi, \tilde{F}^{opt})$, where $\tilde{F}^{opt}$ is the optimal filter for the scan Ψ, we have

$$\left| E_Q L_{(\Psi, \tilde{F}^{opt})}(X_B, Y_B) - \tilde{U}(l_H, Q_B) \right| \leq 2\epsilon_\delta.$$

Example

For $\delta = 0.1$, $\epsilon_\delta < 0.035$, yielding a maximal loss of 0.07.

Filtering is less sensitive to the scanning order

Compare this to 0.16 in the prediction scenario (or even larger values in the noisy prediction scenario).

Excess Loss Bounds - Binary Input and BSC

Theorem

Let Y_t be the output of a BSC(δ) whose input is X_t. Then, for any scanterer $(\Psi, \tilde{F}^{opt})$, where $\tilde{F}^{opt}$ is the optimal filter for the scan Ψ, we have

$$\left| E_Q L_{(\Psi, \tilde{F}^{opt})}(X_B, Y_B) - \tilde{U}(l_H, Q_B) \right| \leq 2\epsilon_\delta.$$

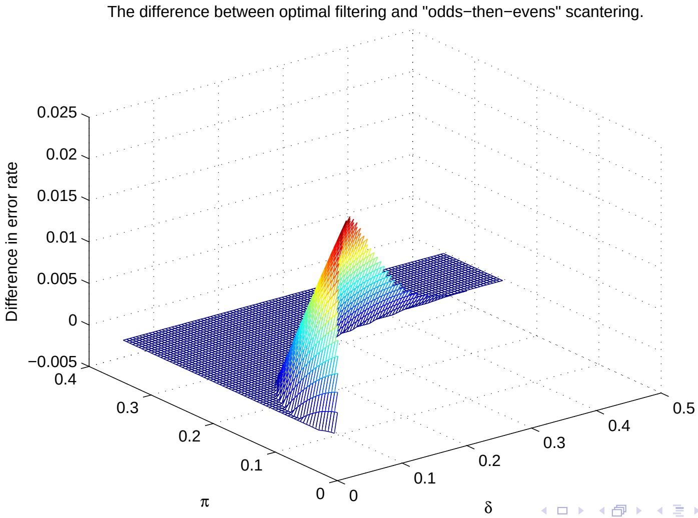
Example

For $\delta = 0.1$, $\epsilon_\delta < 0.035$, yielding a maximal loss of 0.07.

Filtering is less sensitive to the scanning order

Compare this to 0.16 in the prediction scenario (or even larger values in the noisy prediction scenario).

Example - First Order Markov Process



Noisy Scandiction - Best Achievable Performance

Corollary

Let $Y_t = X_t + N_t$ and assume both X and N are Gaussian. The noisy scandictability of Q is then given by

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

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.

Universal (noisy) Scandiction

Theorem

Let X be a stationary random field over a finite alphabet A and a probability measure Q. Let Y be the output of a binary symmetric channel whose input is X and whose crossover probability δ. Then, there exists a sequence of scandictors $\{(\Psi, F)_n\}$, independent of Q and of δ, for which

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

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

Excess Loss Bounds

Corollary (Scan invariance: continuous time mutual information)

Assume Gaussian input under AWGN. Then, for any Ψ^1 and Ψ^2,

$$\frac{1}{|B|}\left|EL_{(\Psi^1,F^{opt})}(X_B,Y_B) - EL_{(\Psi^2,F^{opt})}(X_B,Y_B)\right| \leq \sigma_X^2 \frac{g(SNR)}{SNR},$$

where $g(x) = x - \ln(1+x)$.

Corollary (Scan invariance: discrete time entropy)

Let (X_B, Y_B) be an arbitrarily distributed field. Then, for any scan Ψ,

$$\left|\frac{1}{|B|}E_{Q_B}L^{l}_{\Psi,F^{opt}}(X_B,Y_B) - \bar{U}(l,Q_B)\right| \leq 2\epsilon_\rho,$$

when ρ is such that $E\left\{\rho(Y,F)|\sigma(X)\right\} = l(X,F)$ and ϵ_ρ depends only on ρ. Example: for squared error, $\rho = l_s - \sigma_N^2$, so $\epsilon_\rho = \epsilon_{l_s}$. For Hamming loss, $\rho = \frac{l_H-\delta}{1-2\delta}$, so $\epsilon_\rho = \frac{2\epsilon_{l_H}}{1-2\delta}$.

Open Problems and Conclusions

- Numerous scanning possibilities result in a problem substantially richer than the 1-D analogue.
- The best achievable performance was quantified, and excess loss bounds were given.
- Although the sequential decision maker does not have access to X_B, $\tilde{U}(l, Q)$ and $\bar{U}(l, Q)$ are universally achievable!
- Many open problems:
  1. What is the optimal scan for a given random field?
  2. Stronger universality results for the individual setting.
  3. Tighter lower bounds on $\tilde{U}(l, Q)$.
  4. Identify scenarios where $\tilde{U}(l, Q)$ is known, and the achieving schemes.
  5. In particular: Gaussian fields corrupted by AWGN.

Thank you