Information-Driven Inference in Resource Constrained Environments

MURI Annual Review Meeting

Integrated Fusion, Performance Prediction, and Sensor Management for Automatic Target Exploitation

John W. Fisher

September 14, 2007

Research Foci

- Computable Performance Bounds for Inference in Distributed Systems
- Information Driven Sensor Management under Resource Constraints
  - Approximate DP methods
    - $O(10^{95})$ -> $O(10^{5})$
  - Integer programming/constraint generation methods
    - tractable approach achieves 95% of optimal performance where optimal requires 10^{40} evaluations.
- Multi-modal Data Fusion
  - Learning representations suitable for inference in graphical models
- Inference over Graphical Structures
  - Link analysis
  - Multi-modal data association

Year 1 Accomplishments

- Publications
  - 3 journal publications
    - IEEE Trans SP (2), CVIU
  - IEEE Signal Processing Magazine Article
    - Co-authors include several MURI PIs
  - 11 Conference publications
    - AI Stats, ICASSP, SSP Workshop, SPIE, ASAP, Asilomar
  - 2 Book Chapters

Year 1 Accomplishments

- Technical Exchange (beyond conference presentations)
  - IPAM Meeting on Sensor Networks
  - ARO Workshop on Challenges and Opportunities in Image Understanding
    - Organized by Anuj Srivistava
  - DARPA ISAT EXPOSE Workshop
  - ARO Workshop on Signal and Information Processing
    - Organized by Al Hero
  - MLSP ’07
    - Program Committee
  - SSP ‘07
    - Presented Tutorial on Data Fusion in Sensor Networks at SSP
    - Organized Special Session on Statistical Signal Processing in Sensor Networks at SSP ’07
      - Several MURI PIs participated

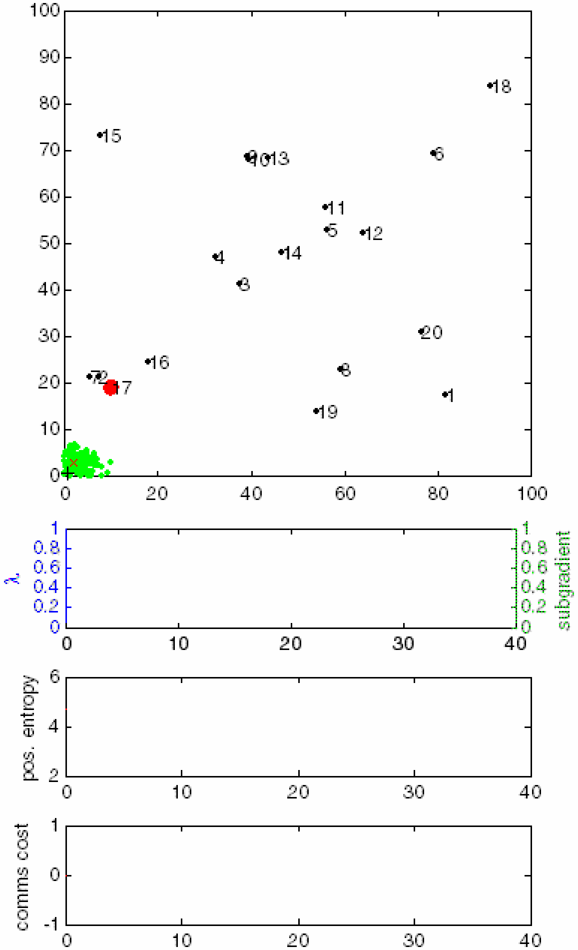
Year 1 Accomplishments

- Completed Ph. D. Students
  - Supervised Jason Williams
    - Willsky, co-supervisor, Castanon, committee member
  - Committee member for Shantanu Joshi
    - Srivastava supervisor

Overview of Technical Presentation

- Performance bounds on information gathering systems

Motivation: Parsimonious Use of Measurements



Intuition

- Measurements are not equally useful and incur different resource expenditures.
- Information regarding many phenomenon is "local".

Zhao, Shin, Reich (2002)

- Consider a tracking application in which sensors yield noisy range measurements.
- Utilize the single sensor measurement which minimizes the expected uncertainty at the next time step.
- Implicitly captures the notion that communications and fusion of all measurements is prohibitive relative to the decrease in uncertainty of the kinematic state.

Motivating Question for Deriving Performance Bounds

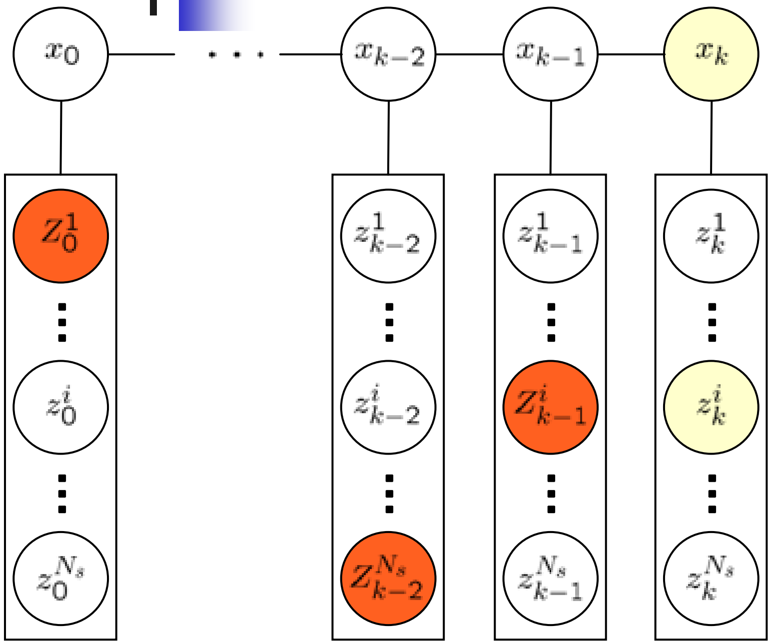
How should we optimize the measurement process for inference problems?

- In many problems acquiring measurements is costly.
- We can select which measurements to obtain in order to perform inference.
- Choices impact both quality of inference and resource expenditures.

Applications

- state estimation/tracking
- identification
- random field estimation

Maximizing Expected Information Gain



Having incorporated previous measurements (or a subset of those available) to compute a posterior

$$p\left(x_k \middle| \{z\}_{0,k-1} = \{Z\}_{0,k-1}\right)$$

choose the sensor whose measurement yields the highest expected information gain.

Notation

z_k^i = measurement of sensor i at time k

Z_k^i = measurement **value** of sensor i at time k

$\{z\}_{i,k}$ = selected measurements from time i to k

$\{Z\}_{i,k}$ = selected measurement **values** from time i to k

Equivalent Information-theoretic Criterion

$$\arg\min_j h\left(x_k|z_k^j, \{z\}_{0,k-1} = \{Z\}_{0,k-1}\right)$$

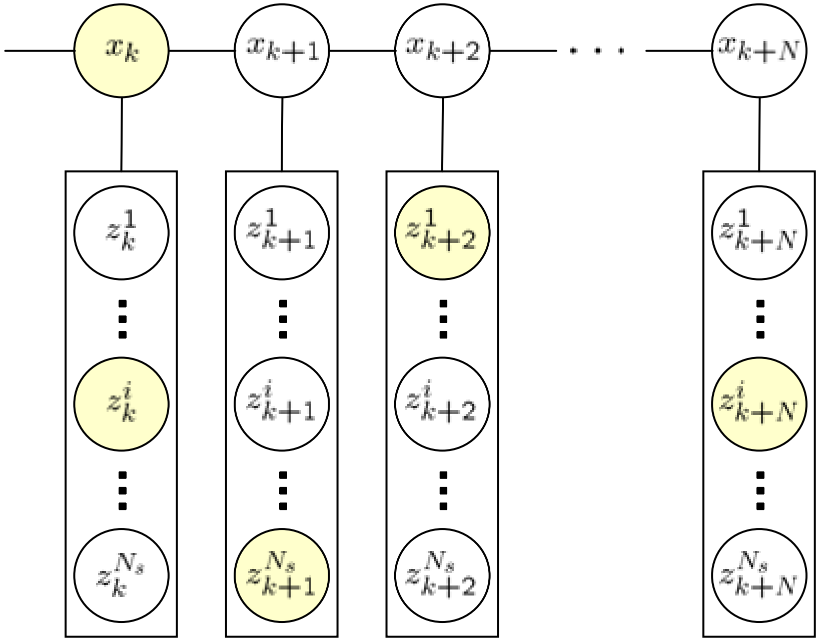
$$\arg\max_j I\left(x_k; z_k^j| \{z\}_{0,k-1} = \{Z\}_{0,k-1}\right)$$

$$\arg\max_j D\left(p\left(x_k|z_k^j, \{z\}_{0,k-1} = \{Z\}_{0,k-1}\right) || p\left(x_k| \{z\}_{0,k-1} = \{Z\}_{0,k-1}\right)\right)$$

$$\arg\min_j D\left(p\left(x_k|z_k^0, \cdots, z_k^{N_s}, \{z\}_{0,k-1} = \{Z\}_{0,k-1}\right) || \; p\left(x_k|z_k^j, \{z\}_{0,k-1} = \{Z\}_{0,k-1}\right)\right)$$

- These are **myopic** (searching over a small number of time steps) and **greedy** (searching over the single best available measurement at each time step).

Value of Long Term Planning?



Choosing the optimal set of measurements (sensors) is exponential in the planning horizon and combinatoric over the number of measurements.

- Can we bound the difference in performance of approximate (tractable) algorithms as compared to optimal?
- Is there a performance gain if we incorporating planning over a longer time-horizon (non-myopic)?
  - Conversely, can we bound the maximal gain of planning over myopic?

Why Use Information-Theoretic Objective Criteria?

Scheffe's Theorem

$$\int |p-q| = 2 \sup_A \left| \int_A p - \int_A q \right|$$
$$= 2 \int_{p>q} (p-q)$$
$$= 2 \int_{q>p} (q-p)$$
$$\left| \int_A p - \int_A q \right| \leq \frac{1}{2} \int |p-q|$$

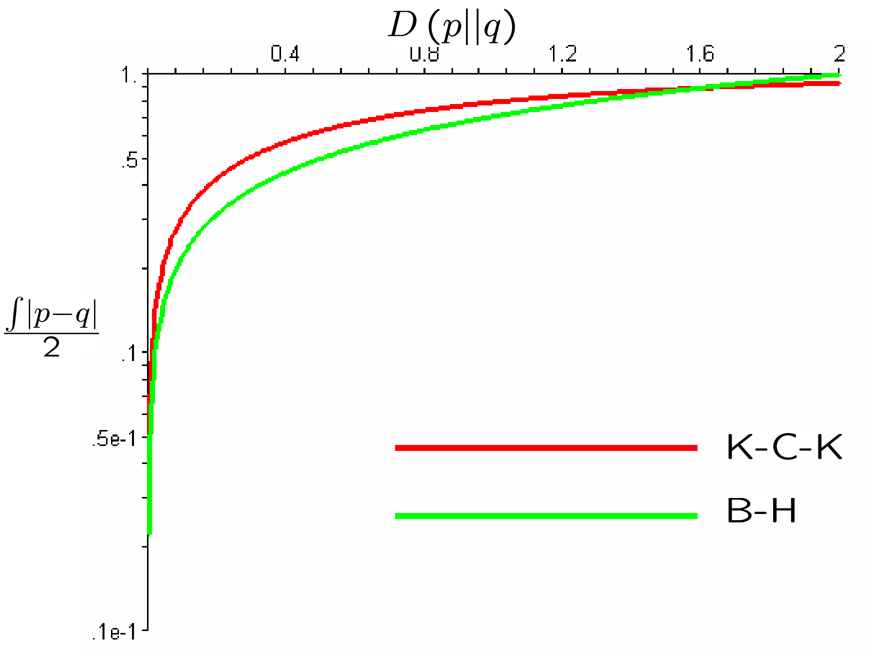
Kullback-Csiszar-Kemperman Inequality

$$\int |p-q| \leq \min \left\{ \sqrt{2D(p||q)}, \sqrt{2D(q||p)} \right\}$$

Bretagnolle-Huber Inequalities

$$\int |p-q| \leq 2\sqrt{1-e^{-D(p||q)}}$$
$$\int |p-q| \leq 2 - e^{-D(p||q)}$$
$$\int \min\{p,q\} \geq \frac{1}{2} e^{-D(p||q)}$$

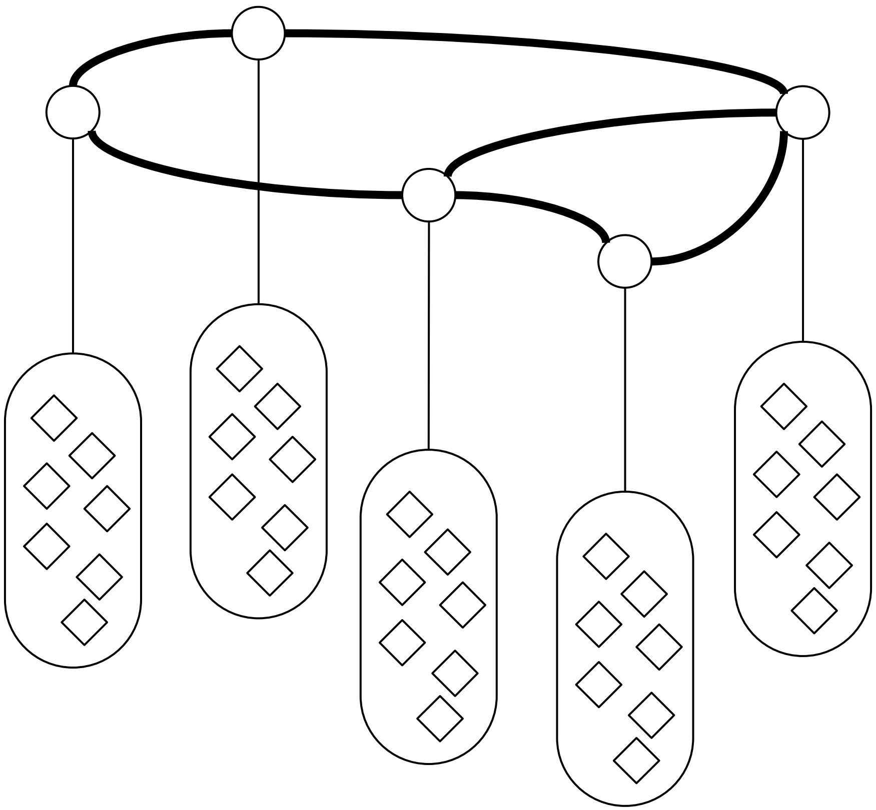
- Closeness in an L_1 sense bounds errors in estimates of event probabilities.
- L_1 is often difficult to optimize, K-L is not in many cases.
- Closeness in K-L bounds closeness in L_1.



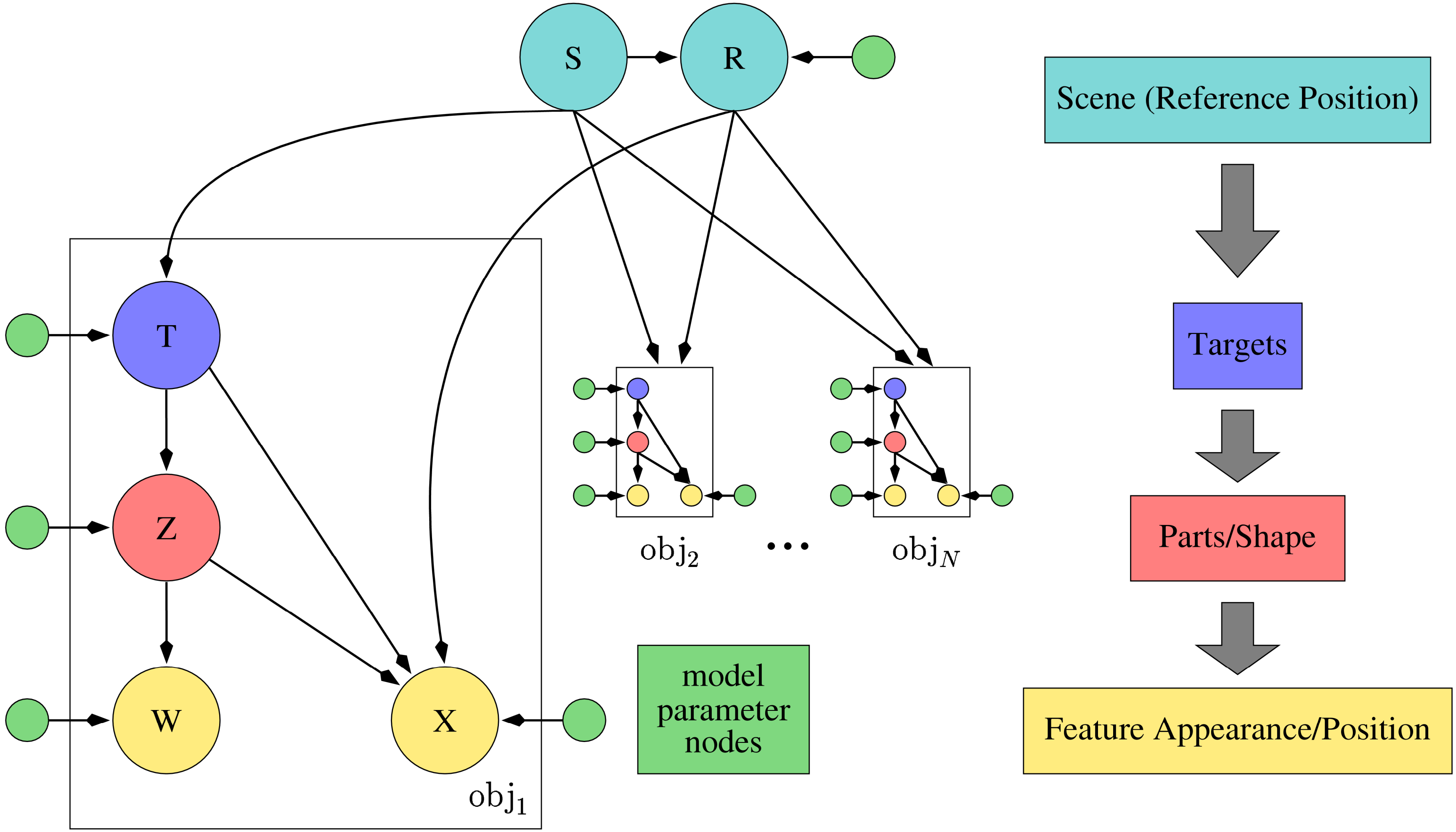
Bounding Competitive Optimal Performance

Optimal schemes are infeasible

- Can we give performance guarantees of approximation schemes as compared to optimal?
- Bound the performance between "greedy" yet tractable versus optimal yet intractable measurement planning algorithms
- Off-line and on-line (typically tighter) bounds



A Complex Scene and Measurement Models



Submodularity

A set function is a real-valued function which takes as its input subsets of a given set: $\mathcal{B}$

$f : \mathcal{A} \in 2^{\mathcal{B}} \to \Re$ where $2^{\mathcal{B}}$ is the set of all subsets of

Def: A set function is *nonnegative* if

$$f(\mathcal{A}) \geq 0 \quad \forall \mathcal{A}$$

Def: A set function is *nondecreasing* if

$$f(\mathcal{B}) \geq f(\mathcal{A}) \quad \forall \mathcal{B} \supseteq \mathcal{A}$$

Def: Denote the *set increment* function as

$$\rho_{\mathcal{B}}(\mathcal{A}) = f(\mathcal{A} \cup \mathcal{B}) - f(\mathcal{A})$$

Def: A set function is *submodular* if

$$f(\mathcal{C} \cup \mathcal{A}) - f(\mathcal{A}) \geq f(\mathcal{C} \cup \mathcal{B}) - f(\mathcal{B}) \quad \forall \mathcal{B} \supseteq \mathcal{A}$$

Capturing Notion of Diminishing Returns

Given sets:

$$\mathcal{A},\mathcal{B},\mathcal{C} \quad \mathcal{A}\subseteq\mathcal{B} \quad \mathcal{B}\cap\mathcal{C}=\emptyset$$

$$\mathcal{B}=\left\{\underbrace{z_1,z_2,z_3,z_4,z_5,z_6}_{\mathcal{A}},\ldots,z_N\right\}$$

$$\mathcal{C}=\{z_{N+1},\ldots\}$$

It is well understood

$$I(X;z^{\mathcal{B}})\geq I(X;z^{\mathcal{A}})$$

<u>Krause and Guestrin (UAI ’05)</u>

Define

$$\begin{aligned}\rho_{\mathcal{C}}(\mathcal{B}) &= I\left(X;z^{\mathcal{B}}\cup z^{\mathcal{C}}\right)-I\left(X;z^{\mathcal{B}}\right)\\ &= I\left(X;z^{\mathcal{C}}|z^{\mathcal{B}}\right)\end{aligned}$$

If observations are independent conditioned on the state, then mutual information is *submodular*

$$\begin{aligned}I(X;z^{\mathcal{C}}|z^{\mathcal{B}}) &= H(z^{\mathcal{C}}|z^{\mathcal{B}})-H(z^{\mathcal{C}}|z^{\mathcal{B}},X)\\ &= H(z^{\mathcal{C}}|z^{\mathcal{B}})-H(z^{\mathcal{C}}|X)\\ &\leq H(z^{\mathcal{C}}|z^{\mathcal{A}})-H(z^{\mathcal{C}}|X)\\ &= I(X;z^{\mathcal{C}}|z^{\mathcal{A}})\end{aligned}$$

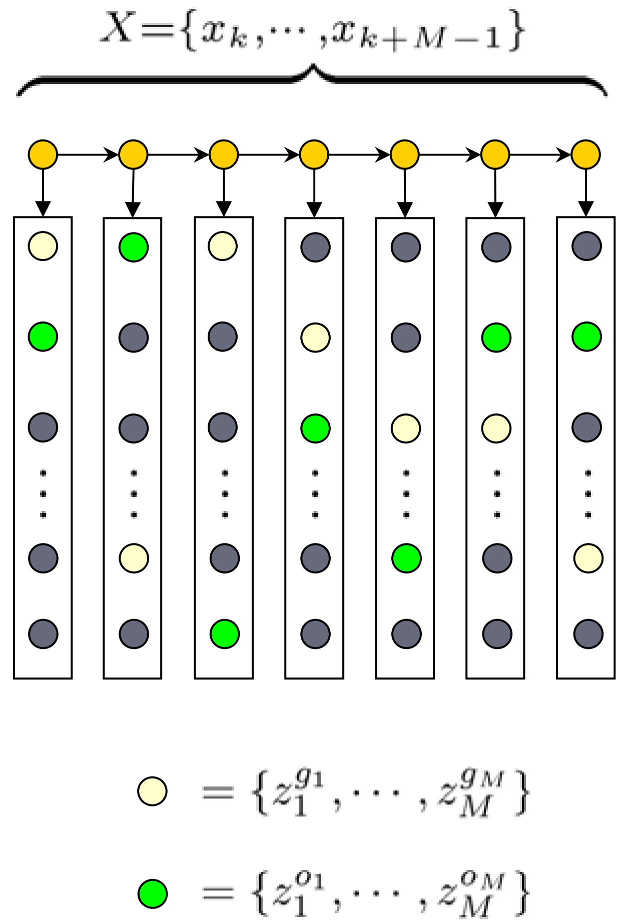
An Open Loop Bound on M-stages

- Previous result is not applicable
- Suppose we have *M* stages
  - Each stage involves selection of an observation for a different sensor or a different time step
- Suppose we use the greedy heuristic to select each observation:

$$g_j = \arg\max_{g \in \{1,\ldots,n_j\}} I(X; z_j^g | z_1^{g_1}, \ldots, z_{j-1}^{g_{j-1}})$$

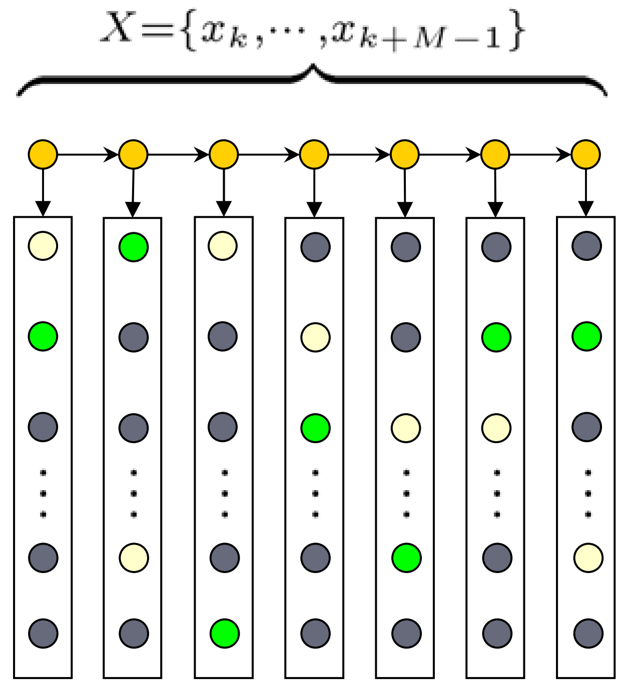
Then…

$$I(X; z_1^{o_1}, \ldots, z_M^{o_M}) \leq 2I(X; z_1^{g_1}, \ldots, z_M^{g_M})$$



An Open Loop Bound

- Bound holds for any path through the sets of measurements.
- Bound holds for arbitrary graphical structure on the latent variable X, not just markov chains.



$$I(X; z_1^{o_1}, \ldots, z_M^{o_M}) \leq 2I(X; z_1^{g_1}, \ldots, z_M^{g_M})$$

(pale yellow circle) $= \{z_1^{g_1}, \cdots, z_M^{g_M}\}$

(green circle) $= \{z_1^{o_1}, \cdots, z_M^{o_M}\}$

The bound is tight

- Consider a simple example where $X = [a \;\; b]^T$, where a and b are independent binary random variables
  - $P(a = 0) = P(a = 1) = 0.5 \;\; [H(a) = 1]$
  - $P(b = 0) = 0.5 - \varepsilon$; $P(b = 1) = 0.5 + \varepsilon \;\; [H(b) = 1 - \delta(\varepsilon)]$
- Available observations are:
  - At first time step, $z_1^1 = a$ or $z_1^2 = b$
  - At second time step, $z_2^1 = a$
- Greedy heuristic chooses z_1^1, z_2^1 for reward 1
- Optimal chooses z_1^2, z_2^1 for reward $2 - \delta(\varepsilon)$

Online computable bound

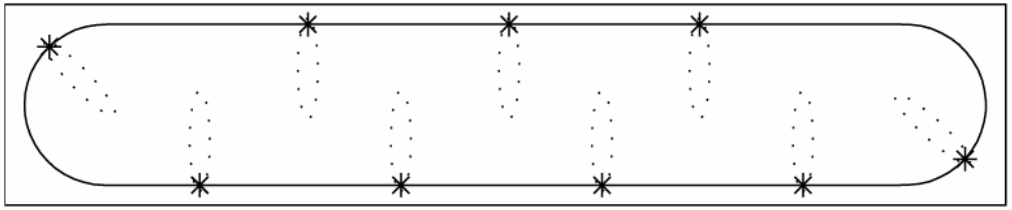
- While the bound is tight, the proof gives rise to an online computable version which may be stronger in particular circumstances
- Let $\bar{g}_i = \underset{\bar{g} \in \{1,\ldots,n_i\}}{\arg\max}\; I(X; z_i^{\bar{g}} | z_1^{g_1}, \ldots, z_M^{g_M})$
- Then:

$$I(X; z_1^{o_1}, \ldots, z_M^{o_M}) \leq I(X; z_1^{g_1}, \ldots, z_M^{g_M}) + \sum_{i=1}^{M} I(X; z_i^{\bar{g}_i} | z_1^{g_1}, \ldots, z_M^{g_M})$$

- This bound is tight in situations where the greedy selection leaves little information behind

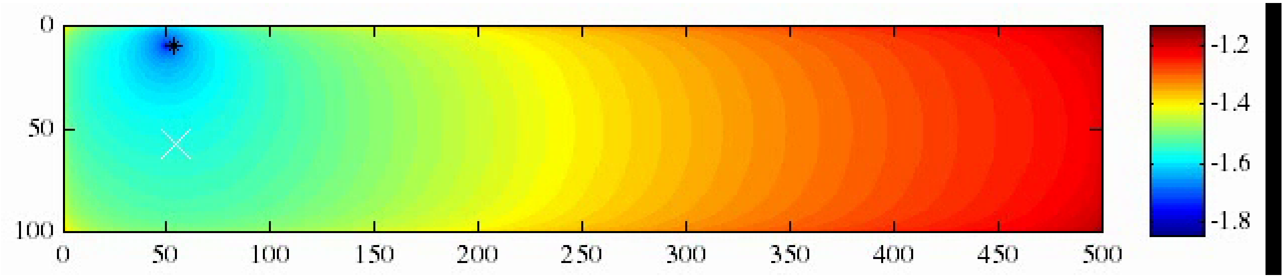
Online computable bound – example

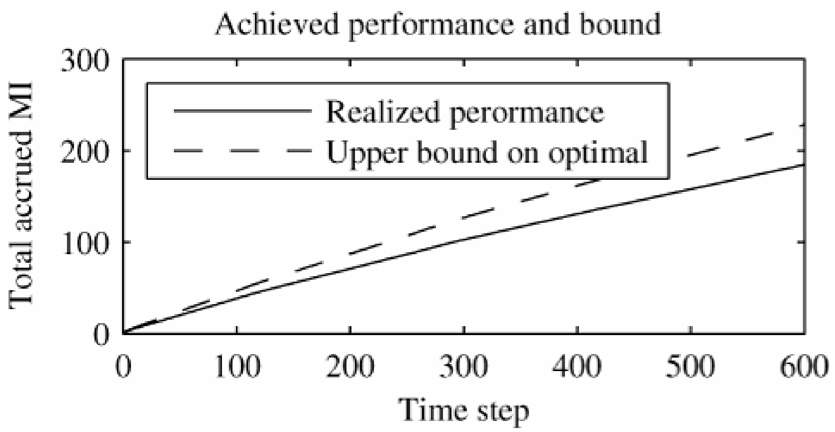
- Suppose we model the depth of the ocean as a Gauss-Markov random field thin membrane model (500 × 100 cells)
- We seek to measure ocean depth using a surface vehicle traveling along a fixed path
  - Available observations depend on current vehicle position

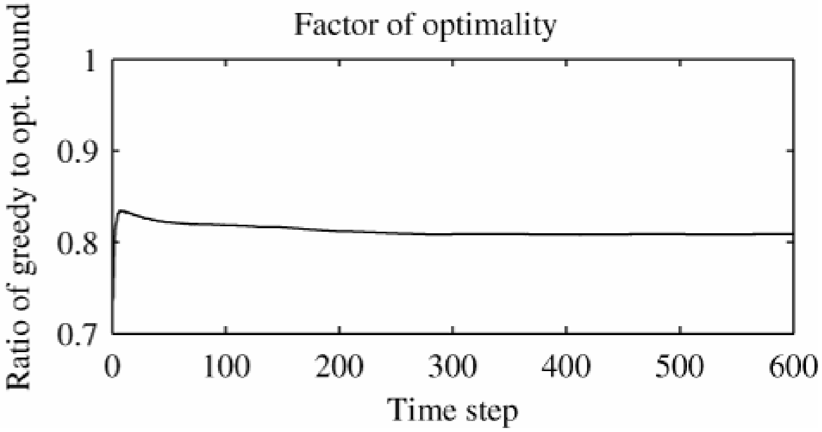


- We need to choose observations at each time

Online computable bound – example





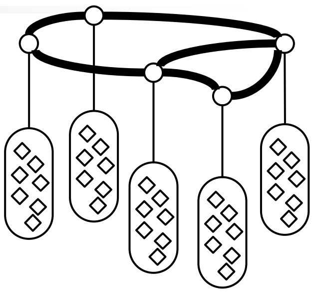


Diffusiveness

- Restrict ourselves to taking one measurement from each set.
  - If the state at one time step is independent of the state at other time steps, the greedy algorithm is optimal
  - One would expect that as dynamics noise increases the same effect would take place
- We can exploit diffusiveness of the dynamics process to achieve a stronger guarantee
- Make use of the following property:

$$I(X; z_j^i | z^{\mathcal{A}}) = I(x_i; z_j^i | z^{\mathcal{A}})$$

Diffusiveness



If there exists an $\alpha \leq 1$ such that for each $i \in \{1,\ldots,L\}$

$$I(x_{\mathcal{N}(w_i)}; z_{w_i}^{j} | z^{\mathcal{G}_i - 1}) \leq \alpha I(x_{w_i}; z_{w_i}^{j} | z^{\mathcal{G}_i - 1})$$

- Then (off-line) the gap between greedy and off-line is reduced

$$I(X; z_{w_1}^{o_1}, \ldots, z_{w_L}^{o_L}) \leq (1+\alpha) I(X; z_{w_1}^{g_1}, \ldots, z_{w_L}^{g_L})$$

- It may be difficult establish the diffusive property everywhere (or over some subset), however one can establish an on-line bound which exploits diffusiveness to the extent that it exists:

$$I(X; z_{w_1}^{o_1}, \ldots, z_{w_L}^{o_L}) \leq I(X; z_{w_1}^{g_1}, \ldots, z_{w_L}^{g_L}) + \sum_{j=1}^{L} I(x_{\mathcal{N}(w_i)}; z_{w_j}^{g_j} | z^{\mathcal{G}_j - 1})$$

- Can be further specialized to trees (i.e. graphs with ordering)

Comments

- While many of the proofs are straightforward the bounds are surprising in 2 ways:
  - It was unexpected that one could bound the performance between an algorithm which ignores the values of future measurements (greedy) and one which considers *all* possible sequences (optimal)
  - Differs from a similar **matroid** guarantee in that one need only consider observations at a single node rather than all available observations at all nodes.
  - It puts more burden on any computationally complex planning algorithm in order to justify its use.
- Bounds are applicable to a variety of information rewards:
  - E.g. the posterior CRB
- Can be extended to the case where the reward (e.g. MI) is approximated with bounded error.
- Can be extended for the case where rewards are discounted over time to get tighter bounds
  - Related to minimum time estimation problems