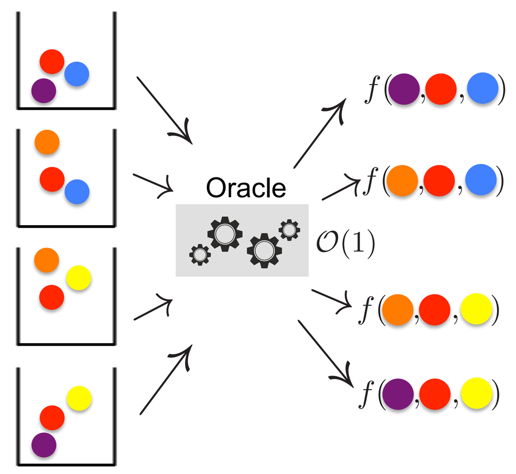
On the Complexity of Information Planning in Gaussian Models

Georgios Papachristoudis and John W. Fisher III
geopapa@mit.edu fisher@csail.mit.edu
Massachusetts Institute of Technology

Motivation

- [2] Information planning considers the problem of finding the most informative subset of measurements for some inference problem.
- Optimal selection is generally intractable as the number of latent and measurement nodes increases.
- For submodular problems, approximate techniques provide nearly optimal solutions.
- These techniques are usually one-step greedy approaches that select the best measurement at each step given past selections.
- Prior analyses often assume an *oracle-value* model; i.e., the reward of any set is computable in $\mathcal{O}(1)$ time.
- In many settings, the *oracle-value* model does not hold.
- Additionally, the reward varies depending on the order in which subsets are considered.
- Consequently, we consider efficient methods for computing information rewards over multiple orderings.



Contributions

- We derive an approach that leads to substantial complexity reduction for information planning in Gaussian models under the assumption that a measurement depends only on a few latent variables.
- We demonstrate experimentally the advantages of this approach as the number of measurements and sparsity increases.
- We derive a related variant of belief propagation which is better suited for information planning settings.
- We show empirically that the benefits of the latter method are more significant as the distance between consecutive elements of the visitation order decreases.

Related work

- [?] consider the problem of selecting an optimal k-element subset of measurements from a ground set $\mathcal{V}$ of size N that maximizes some reward f.
  - They provide a $(1 - 1/e)$ worst-case bound.
- [?] provide a 50% lower bound to matroidal structures .
- [?, ?] consider information planning in the batch setting, where all measurements are available.
  - They show that entropy and mutual information (under mild conditions) are submodular.
- [3] consider influential subset selection in social networks.
- Future work builds upon or extends the work of Nemhauser et al. (1978). I.e., work by [1] on stochastic submodular maximization, by [5, 4] on online resource allocation networks and by [6] on the submodular welfare problem.

- All previous works assume the *oracle value* model; a universal "oracle" provides the function value for any input set in constant time.
- Evaluation of rewards usually presents a significant computational challenge.
  - [?] propose truncation methods in Gaussian models.
  - [?] note that the evaluation of entropies can be prohibitive.
  - [3] acknowledge the complexity of evaluating the influence function.

Gaussian HMMs

- We restrict ourselves to HMMs (results easily extended to trees and polytrees).

$$\mathrm{X}_t = A_{t-1}\mathrm{X}_{t-1} + \mathrm{V}_{t-1} \qquad \mathrm{Y}_t = C_t\mathrm{X}_t + \mathrm{W}_t,$$

where $\mathrm{V}_t \sim \mathcal{N}(0, Q_t), \mathrm{W}_t \sim \mathcal{N}(0, R_t)$ and A_t, C_t are the dynamics and measurement matrices, respectively.

Problem statement

Notation

- $\boldsymbol{X} = \{\mathrm{X}_1, \ldots, \mathrm{X}_T\}$: T latent variables that are the focus of inference ($\mathrm{X}_t \in \mathbb{R}^d$).
- $\{\mathcal{V}_1, \ldots, \mathcal{V}_T\}$: T observation sets ($|\mathcal{V}_t| = N_t$).
- $\mathbf{Y} = \{\mathrm{Y}_1, \ldots, \mathrm{Y}_T\}$: Each Y_t is an $N_t m \times 1$ vector holding the m measurements of set $\mathcal{V}_t$ ($\mathrm{Y}_{t,u} \in \mathbb{R}^m$).

Goal

- Select up to k_t measurements from each set $\mathcal{V}_t$ that maximize some reward of interest

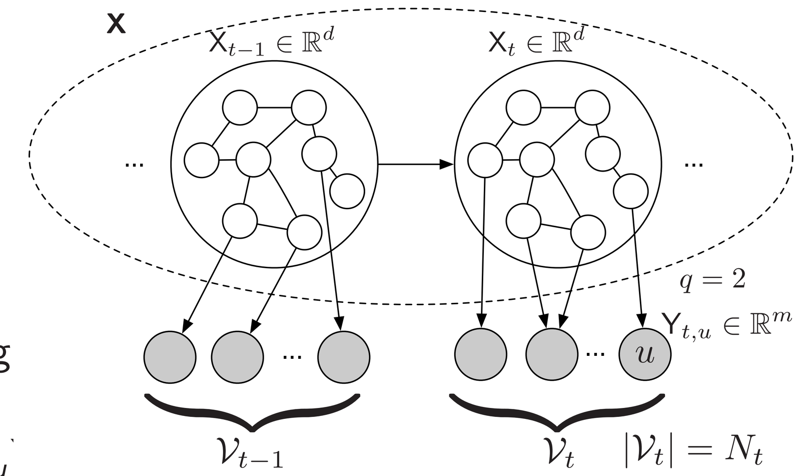
$$\mathcal{O} \in \underset{|\mathcal{S}_t| \le k_t}{\arg\max}\, f(\mathcal{S}), \text{ where } \mathcal{S} = \bigcup_{k=1}^{T} \mathcal{S}_k,\ \mathcal{S}_i \cap \mathcal{S}_j = \emptyset, \forall i \neq j. \quad (1)$$

Reward

- Reward function: $f : 2^{\mathcal{V}} \mapsto \mathbb{R}$ / Incremental reward: $f(u \mid \mathcal{S}) \triangleq f(\{u\} \cup \mathcal{S}) - f(\mathcal{S})$.
- Monotone: $f(\mathcal{A}) \le f(\mathcal{B}), \forall \mathcal{A} \subseteq \mathcal{B} \Leftrightarrow f(u \mid \mathcal{A}) \ge 0, \forall u \in \mathcal{V} \setminus \mathcal{A}, \mathcal{A} \subseteq \mathcal{V}$.
- Submodular: $f(u \mid \mathcal{A}) \ge f(u \mid \mathcal{B}), \forall u \in \mathcal{V} \setminus \mathcal{B}, \mathcal{A} \subseteq \mathcal{B}$ [?].

Sparsity

- Measurements are obtained from a small subset of the underlying process $\boldsymbol{X}$ $\Rightarrow$ Sparsity of measurement matrix C_t.
  - Each m-row block of C_t (corresponding to a measurement $\mathrm{Y}_{t,u}$) has at most q non-zero entries.



Greedy selection and theoretical guarantees

For a specified visitation order $\{w_1, \ldots, w_M\}$ such that $\sum_{j=1}^{M} \mathbb{1}(w_j = t) = k_t$, select g_j that maximizes

$$g_j = \underset{u \in \mathcal{V}_{w_j} \setminus \mathcal{G}_{j-1}}{\arg\max}\, f(u \mid \mathcal{G}_{j-1}). \quad (2)$$

Bound [?] $f(\mathcal{O}) \le 2f(\mathcal{G})$ for problem (1), when f monotone and submodular. (3)

- For every step in (2), we need to compute the incremental reward of all remaining measurements in $\mathcal{V}_{w_j}$.

Mutual Information (MI)

- $f(\mathcal{S}) = I(\boldsymbol{X}; \mathrm{Y}_{\mathcal{S}}) \quad \Rightarrow \quad f(u \mid \mathcal{S}) = I(\boldsymbol{X}; \mathrm{Y}_u \mid \mathrm{Y}_{\mathcal{S}})$.
- **Monotone**: $I(\boldsymbol{X}; \mathrm{Y}_{\mathcal{S}}) \ge 0$ & **submodular**: $I(\boldsymbol{X}; \mathrm{Y}_u \mid \mathrm{Y}_{\mathcal{A}}) \ge I(\boldsymbol{X}; \mathrm{Y}_u \mid \mathrm{Y}_{\mathcal{B}}), \forall \mathcal{A} \subseteq \mathcal{B}$, under $\mathrm{Y}_i \perp\!\!\!\perp \mathrm{Y}_j \mid \boldsymbol{X}$.
- Closed-form expression for Gaussian setting:

$$g_j = \underset{u \in \mathcal{V}_{w_j} \setminus \mathcal{G}_{j-1}}{\arg\max}\, f(u \mid \mathcal{G}_{j-1}) = I(\boldsymbol{X}; \mathrm{Y}_u \mid \mathrm{Y}_{\mathcal{G}_{j-1}}) = I(\mathrm{X}_{w_j}; \mathrm{Y}_u \mid \mathrm{Y}_{\mathcal{G}_{j-1}}) \propto \log \frac{|J_{w_j|\{u\} \cup \mathcal{G}_{j-1}}|}{|J_{w_j|\mathcal{G}_{j-1}}|} \Rightarrow \text{Dependence only on } \mathrm{X}_{w_j}.$$

  - Ratio of uncertainties (as expressed by inv covariance) after and before the incorporation of measurement u.

Complexity of evaluating rewards not assuming sparsity

- **Update**: Update the covariance at the current point after the incorporation of measurement g_j.

$$\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1} C_t^T (C_t \Sigma_{t|t-1} C_t^T + R_t)^{-1} C_t \Sigma_{t|t-1}.$$

- **Exploration**: Explore all remaining measurements in the current observation set.
- **Complexity**: Complexity per update: $\mathcal{O}(md^2)$ / Complexity per exploration: $\mathcal{O}(Nd^3)$.

Complexity Reduction due to Sparsity

Each measurement is drawn at most from q latent variables:
I_c denotes the indicator matrix; 1 where C_t is non zero.

Reductions during update

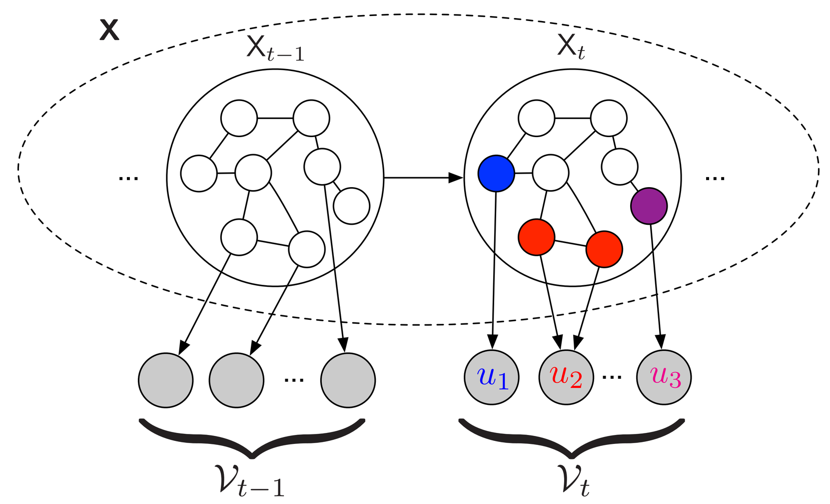
For a chosen measurement u, update the covariance as

$$\Sigma' = \Sigma - \Sigma C(u,:)(C(u,:)\Sigma C(u,:)^T + R(u,u))^{-1} C(u,:)\Sigma$$

$$J' = J + C(u,:)^T R(u,u)^{-1} C(u,:), \quad (4)$$

$$J' = J + \begin{bmatrix} \hat{C}_u^T \\ 0^T \end{bmatrix} \begin{bmatrix} \hat{C}_u & 0 \end{bmatrix} \Rightarrow J(I_u, I_u)' = J(I_u, I_u) + \hat{C}_u^T \hat{C}_u, \quad (5)$$

where $J = \Sigma^{-1}$, $\hat{C}_u = R(u,u)^{-1/2} C(u, I_u)$, $I_u = I_c(u,:)$.



Complexity Reduction due to Sparsity (cont.)

For step j: $J' = J'_{w_j}$, $J = J_{w_j}$, $C = C_{w_j}$, $R = R_{w_j}$.

- Complexity per update: $\mathcal{O}(m \max\{m, q\}^2)$.
- It affects only the block of J directly linked to measurement u: $\mathcal{O}((\frac{d}{\max\{m,q\}})^2)$.

Reductions during exploration

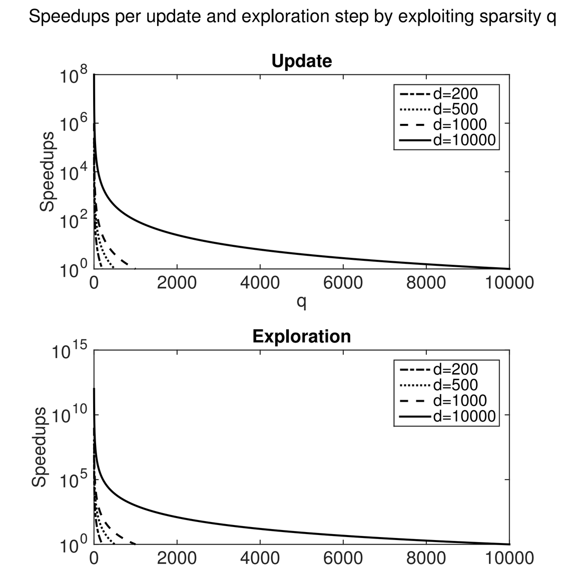
Compute the incremental value of all measurements and select the one with the highest reward:

$$g_j = \underset{u \in \mathcal{V}_{w_j} \setminus \mathcal{G}_{j-1}}{\arg\max}\, I(\mathrm{X}_{w_j}; \mathrm{Y}_{w_j,u} \mid \mathrm{Y}_{\mathcal{G}_{j-1}}) = I(\mathrm{X}_{w_j,\mathcal{N}(u)}; \mathrm{Y}_{w_j,u} \mid \mathrm{Y}_{\mathcal{G}_{j-1}})$$

$$= \log \frac{|J_{w_j|\{u\} \cup \mathcal{G}_{j-1}}(I_u, I_u)|}{|J_{w_j|\mathcal{G}_{j-1}}(I_u, I_u)|} \quad (6)$$

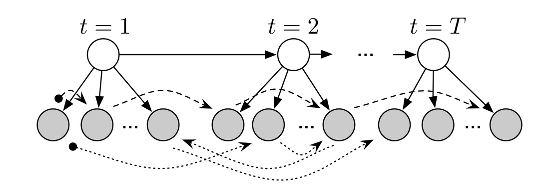
$$\overset{(5)}{=} \log |\mathbb{I}_{m \times m} + \hat{C}_{w_j,u} \Sigma_{w_j|\mathcal{G}_{j-1}}(I_u, I_u) \hat{C}_{w_j,u}^T|. \quad (7)$$

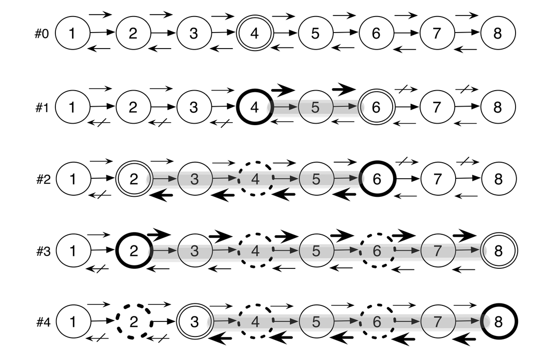
- Complexity per exploration: $\mathcal{O}(Nm \max\{m, q\}^2)$.
- Only block of J directly linked to measurement u is relevant: $\mathcal{O}((\frac{d}{\max\{m,q\}})^3)$.



Incremental Belief Propagation (IncrBP)

- $\binom{\sum_t k_t}{k_1, \ldots, k_T} = \frac{(kT)!}{k!^T}$ walks (visitation orders).
- Different walks can result in very different rewards.
  - High-reward walks can serve as good solutions.
  - Low-reward walks as tighter upper bounds for the optimal solution (see Eq. (3)).
- For a walk $\{w_1, \ldots, w_M\}$, when we move from $w_j \to w_{j+1}$, we only update $\Sigma_{w_{j+1}|\mathcal{G}_j}$.
  - Update node potential at X_{w_j}.
  - If $w_j < w_{j+1}$: Propagate messages from w_j to w_{j+1}.
  - If $w_j > w_{j+1}$: Propagate messages from w_{j+1} to w_j.





References

- A. Asadpour, H. Nazerzadeh, and A. Saberi. Stochastic Submodular Maximization. *In Proceedings of the 4th International Workshop on Internet and Network Economics, WINE*, pages 477–489, December 2008.
- Peter L. Bartlett, Michael I. Jordan, and Jon D. Mcauliffe. Convexity, classification, and risk bounds. *Journal of the American Statistical Association*, 2003.
- D. Kempe, J. Kleinberg, and É. Tardos. Maximizing the Spread of Influence Through a Social Network. *In Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '03, pages 137–146, New York, NY, USA, 2003. ACM.
- M. Streeter, D. Golovin, and A. Krause. Online Learning of Assignments. *In Advances in Neural Information Processing Systems 22*, pages 1794–1802, 2009.
- M. J. Streeter and D. Golovin. An Online Algorithm for Maximizing Submodular Functions. *In Proceedings of the Twenty-Second Annual Conference on Neural Information Processing Systems*, pages 1577–1584, December 2008.
- J. Vondrák. Optimal Approximation for the Submodular Welfare Problem in the Value Oracle Model. *In Proceedings of the Fortieth Annual ACM Symposium on Theory of Computing*, STOC '08, pages 67–74, New York, NY, USA, 2008. ACM.