Tracking Time-varying Graphical Structure

Erich Kummerfeld & David Danks

Dept. of Philosophy, Carnegie Mellon University

The Challenges

- Standard causal learning methods assume structure is stable through time
- In fast-paced settings, we cannot wait until data collection is finished to do causal learning
- Memory constraints can prevent storing all data

Desired Algorithm

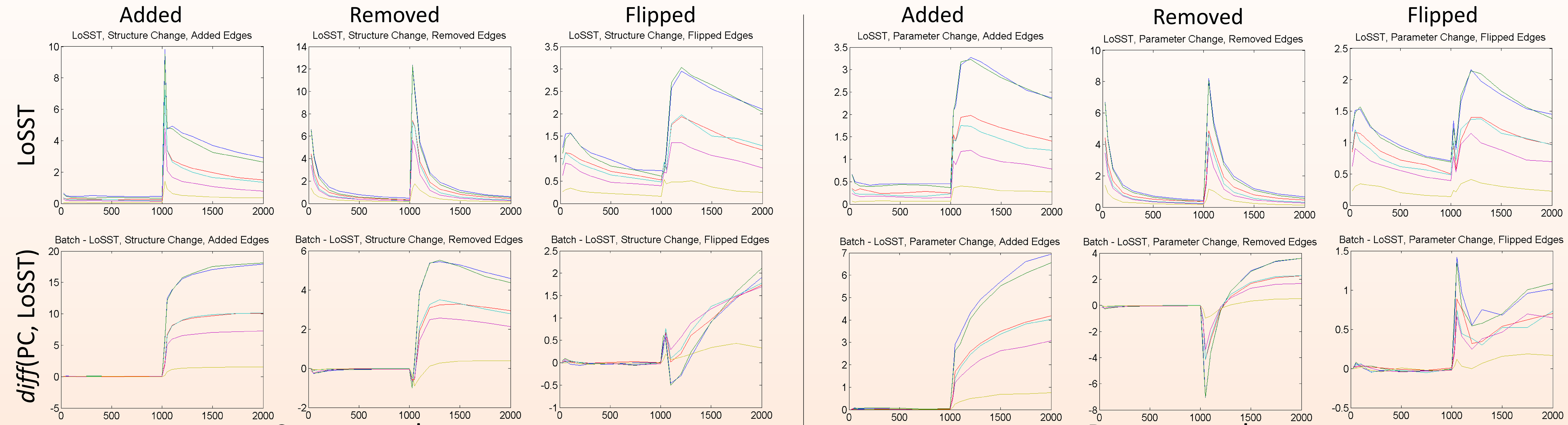
- Method that:

(A) can learn in "real-time" (i.e., without storing all data & running only at the end); and

(B) can "track" possibly-changing underlying causal structure

Simulations:

All graphs show average edge errors



Structure change:

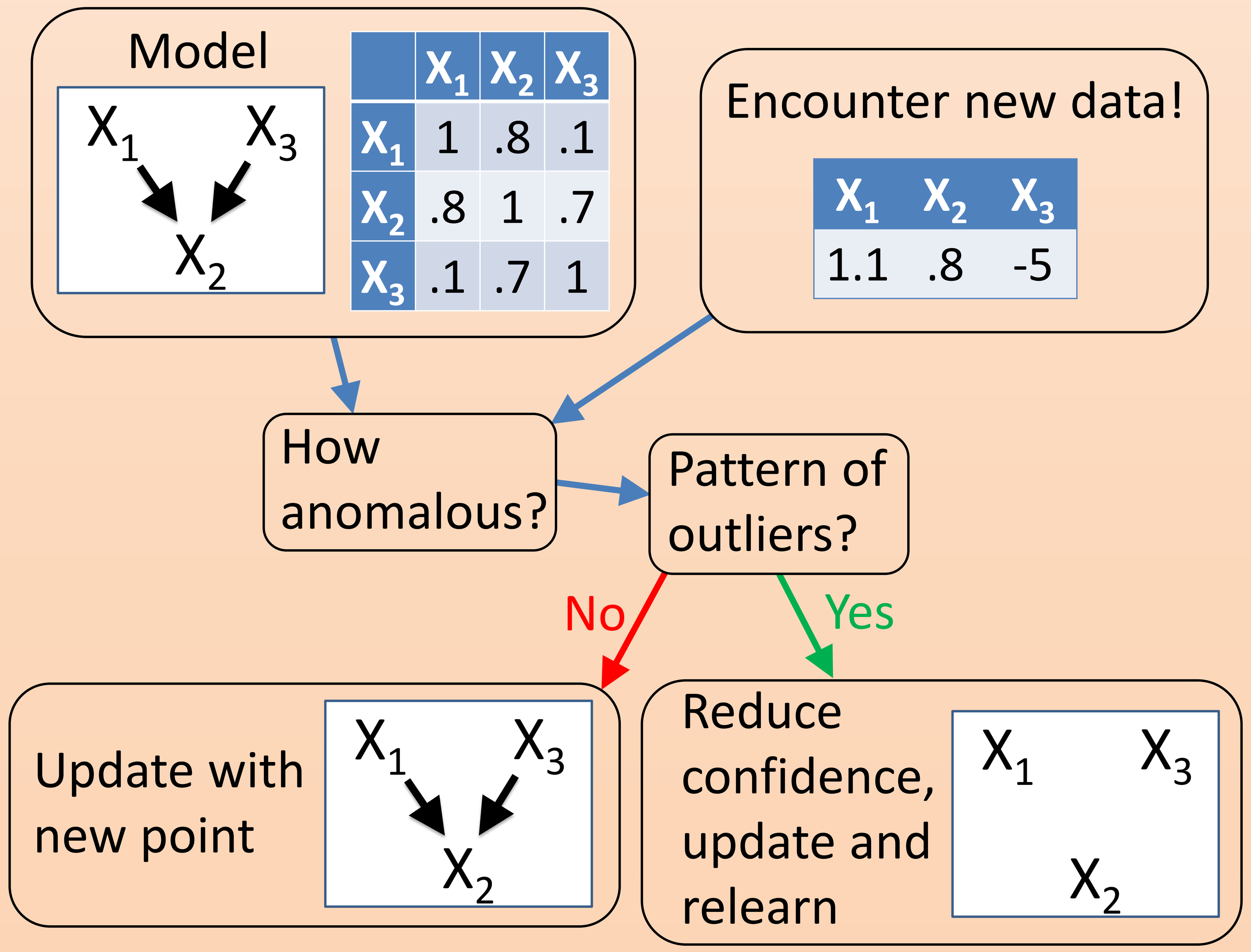
- LoSST responds rapidly to structure change
- LoSST shadows PC before change, outperforms PC after change
- Major difference: PC adds many more edges than LoSST after change

Parameter change:

- Fewer errors overall with parameter changes
- LoSST outperforms PC; PC is briefly "lucky" post-change on edges removed
- PC continues to accumulates more false edges over time

LoSST

(***Lo***cally ***S***tationary ***S***tructure ***T***racker)



LoSST Module Details

Sum of weights: $b_r = \sum_{k=1}^{r} a_k$

Weighted mean: $\mu_i^r = \sum_{k=1}^{r} \frac{a_k}{b_r} X_i^k$

Updating the mean: $\mu_i^{r+1} = \frac{b_r}{b_{r+1}} \mu_i^r + \frac{a_{r+1}}{b_{r+1}} X_i^{r+1}$

Weighted covariance: $\mathbf{C}_{V_i,V_j}^r = \sum_{k=1}^{r} \frac{a_k}{b_r} (X_i^r - \mu_i^r)(X_j^r - \mu_j^r)$

Updating the Weighted covariance:

$$\mathbf{C}_{X_i,X_j}^{r+1} = \frac{1}{b_{r+1}} [b_r \mathbf{C}_{X_i,X_j}^r + b_r \delta_i \delta_j + a_{r+1}(X_i^{r+1} - \mu_i^{r+1})(X_j^{r+1} - \mu_j^{r+1})]$$

Correction term: $\delta_i = \mu_i^{r+1} - \mu_i^r = \frac{a_{r+1}}{b_{r+1}} (X_i^{r+1} - \mu_i^r)$

Mahalanobis distance: $D_r = (\mathbf{X}^r - \vec{\mu})(\mathbf{C}^r)^{-1}(\mathbf{X}^r - \vec{\mu})^T$

p-value of the distance: $T^2(x > D_r | p = N, m = S_r - 1)$

Weighted pooled p-value: $\rho_r = \Phi(\sum_{i=1}^{r} a_i \Phi^{-1}(p_i, 1), \sqrt{\sum a_i^2})$

Weight for next data point: $a_{r+1} = \begin{cases} a_r & \text{if } \rho_r \geq T \\ \frac{a_r \gamma T}{\gamma T + \rho_r - T} & \text{otherwise} \end{cases}$

Morals & Lessons

Advantages of LoSST:

- Performs well on both stable and varying structures
- Memory constraints depend on variables, not data
- Can be run with real-time, streaming data or on a single pass through the dataset
- "Plug-and-play" design allows for easy modification to use alternative algorithms
- Simple modification is pointwise convergent

Constraints on LoSST:

- Currently assumes linearity and Gaussianity
- Cannot always track *very* rapid structure change
- Cannot be both consistent and diligent (but provably, no method can)