Batched optimal experimental design for topology reconstruction of Bayesian networks

Joel Eliason[1,*]

[1]Northwestern University, Engineering Sciences and Applied Mathematics, Evanston, IL, USA
[*]joeleliason2021@u.northwestern.edu

ABSTRACT

In this work, I examine batched optimal experimental design (OED) as it is applied to the reconstruction of continuous Bayes network topology as a model for gene regulatory networks (GRNs). In particular, I extended a Bayesian OED method for GRN reconstruction[1] and utilized the ranking of interventional experiments induced by this method to batch experiments. I utilized this "naive" batching strategy in order to yield greater parallelization of experiments and thus potentially massive increases in efficiency. During this investigation, I performed a replication study of[1] and found that their method also performs worse than random at large network sizes (size > 14), though it performs better than random at small network sizes (size < 14). Furthermore, I found the efficacy of such a naive approach to be worse than selection of interventions at random. I tested the hypothesis that correlations between experiments are the mechanism behind the suboptimal performance of batching and found that the addition of the most highly ranked experiment does not change network recovery outcomes. Lastly, I tested the effects of increased noise on active network recovery. Potential mechanisms behind poor recovery for large size are discussed, as well as possible future experiments to predict and counteract correlation between ranked experiments.

Introduction and Background

Data is expensive, especially in the biological sciences. Furthermore, with the dizzying array of potential gene knockdown or other perturbative experiments for the inference of genetic regulatory networks, a researcher can be completely overwhelmed with the choice of experiment to perform. Many researchers, in this case, opt for a systematic strategy of perturbative experiments. However, in some cases, these strategies can be wildly inefficient, as systematic experiments can deliver quite uninformative results, in light of what is already known. Furthermore, all researchers are under a cost constraint and many are unable to systematically perform all relevant perturbations to a GRN with the budget at their disposal. Because of this, there have been several publications since the early 2000s focusing on the optimal experimental design (OED) problem, an attempt to predict the most informative data points, as it relates to network reconstruction[2–6]. However, in comparison with the number of network inference methods that have been developed in the same time, OED has received only a fraction of the attention. It bears worth mentioning that, as OED methods have been developed, they can result in remarkable efficiency and speed-ups of studies[7]. Broadly speaking, many studies of OED have been focused on designing experiments for better parameter calibration. However, OED for topology reconstruction (or model selection, in other areas of statistics) can be a harder problem. This is in part due to the discreteness of models (ie, an integer number of variables), as well as the inherently limited knowledge on how perturbations will propagate through the network.

As it stands, there currently exist several competing methods for OED for topology reconstruction[1–6]. Many of these methods depend on a calculation of the entropy (from information theory), the Kullback-Leibler (KL) divergence, difference in predicted outcomes or optimization of variance. Critically, most of these methods work in a strictly sequential manner: the next, and only the next, experiment is chosen based on the results of all previous experiments. Though there are some methods within the realm of network reconstruction that propose batches of experiments to be run in parallel, the area has received far less attention than even the general OED problem[3,7,8]. In particular, as far as the author can tell, there are no batched OED methods for the inference of continuous Bayesian networks. From the researcher's standpoint, the ability to parallelize experiments can lead to massive increases in efficiency and is thus a very relevant question. As mentioned before, most developed methods only work sequentially. However, many methods, in the process of optimizing a specific metric, induce a ranking of interventional experiments. Therefore, the driving question in this work is: instead of only using the most optimal experiment, what if we used the top N experiments (as predicted by the OED method)? How much confidence do we have in this "naive" batched OED method for predicting multiple experiments? Here, I provide computation-based evidence that such a naive batched OED method performs worse than the random selection of interventions. Furthermore, I show that the method proposed in[1] performs worse than random for network sizes greater than 14, at least at low noise levels.

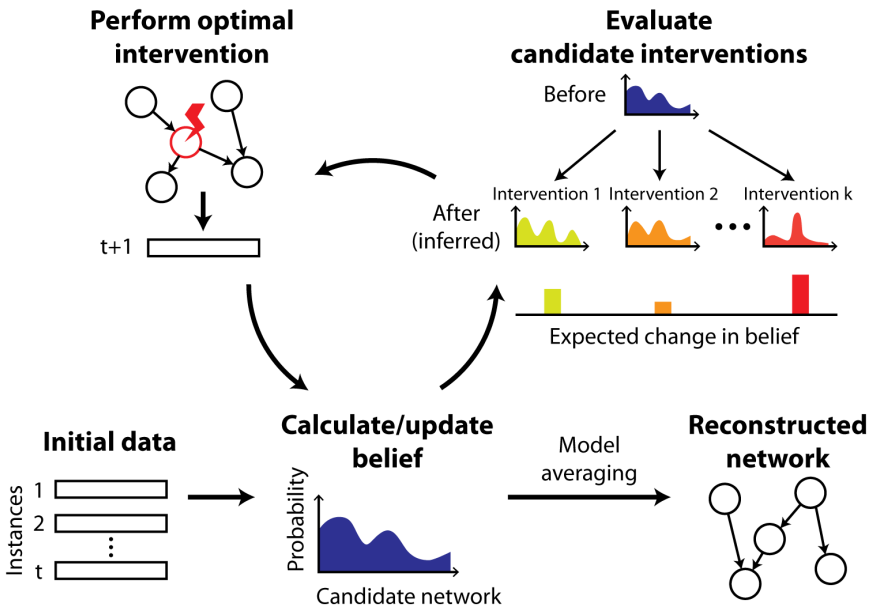


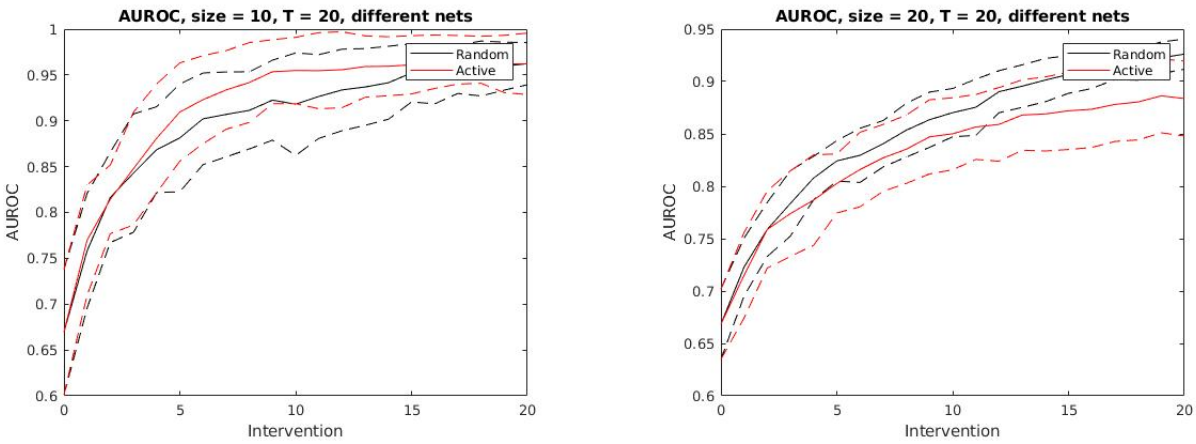
Figure 1. Schematic of the active network recovery process. Note that in the batched recovery process, instead of taking only the most informative experiment (i.e., the "red" experiment in the top right), we would take the N top experiments (e.g. the "red" and "yellow" experiments if N=2). Figure used from[1].

Methods

In this work, I utilized the information-theoretic active learning method for Gaussian Bayesian networks (GBNs) as described in[1]. In GBNs, each node takes on continuous values and has some additive Gaussian noise. This assumption of Gaussanity yields an analytical marginal likelihood that greatly reduces the computational time for Bayesian inference. This OED method is one of the most recent for Bayesian networks (and the first for using Bayesian inference with GBNs), and aims to utilize interventional experiments to not only learn the directionality of edges, as in[3,6], but the underlying skeleton of the topology as well. The metric used to order interventional experiments is the Kullback-Leibler (KL) divergence, a measure of the difference between two probability distributions. Thus, the experiment that maximizes the expected change in belief over candidate graph structures is the interventional experiment that is utilized in the typical active learning strategy. As mentioned in the Background, I used the ranking induced by the KL divergence to batch experiments. See Figure 1 for a schematic of the active learning process. Furthermore, the networks that are used for computational experiments are directed acyclic graphs (DAGs), containing no feedback loops. This additionally simplifies the learning process, as recovery of feedback can be very difficult. Fortuitously, the researchers in[1] have posted a Supplementary Information section that contains MATLAB code for performing OED alongside Bayesian inference, which I used to perform both sequential and batched OED.

I performed 4 different computational experiments during this project. Their results are described below - here, I describe each of the experiments in some detail. Each of these computational experiments tested either the original active learning recovery method or my altered recovery method. I tested this recovery method on randomly generated DAGs of size 10-20 and with a density of 0.4 edges (ratio of edges to non-edges for any given network). All experiments consisted of a comparison of some variant of the active learning strategy (in which experiments were selected according to the KL metric) and the random selection strategy. The randomly generated DAGs had binary edge weights (either on or off), node states were selected from a standard normal distribution, and the variance of each node (used for drawing samples from that node after intervention) was set to 0.0025 for all but the last experiment.

Each computational experiment examined the AUROC, AUPRC and MSE of edges after each intervention, for up to $T = 20$ interventions. Briefly, the AUROC and AUPRC are representative of the ability of a classifier to distinguish between edges and non-edges. In particular, the AUROC is the area under the ROC curve and is a summary of the ability of a classifier to distinguish between true positives and false positives (i.e., how often it labels edges as 'edges' vs how often it labels non-edges as 'edges') and is the result that I will be presenting throughout this report. Therefore, the higher the AUROC curve, the closer the recovery method has come to discovering the true topology of the network. Figures of the AUPRC and MSE will be included in the Supplementary Information as well.



(a) Size = 10 **(b)** Size = 20

Figure 2. Replication study performed with 5 trials each, attempting to recover a different network during each trial. Note that the OED method performs well for size = 10, but performs *worse than random* for size = 20.

Results

Experiment 1: Replication

The first set of experiments that I ran were a replication of[1], in order to test whether or not I could recover random DAGs using the original OED method as well as reported in[1]. I performed these experiments over a network size varying from 10-20, performing recovery on the network of each size 5 times in order to establish an average AUROC, AUPRC or MSE value, as well as a standard deviation around this average value (this standard deviation is indicated by dotted lines around each average). As can be seen in Figure 2, the method as given yields good results for size = 10; however, at size = 20, it loses out to random selection of experiments. Size = 14 is where the active learning strategy starts to lose out to random selection of interventions (see Figure 3). As only nets of size 10 or less were recovered in[1], this was an unforeseen problem and will be commented on further in the Discussion.

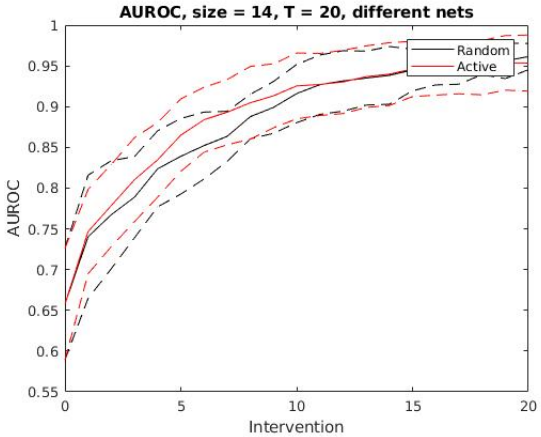
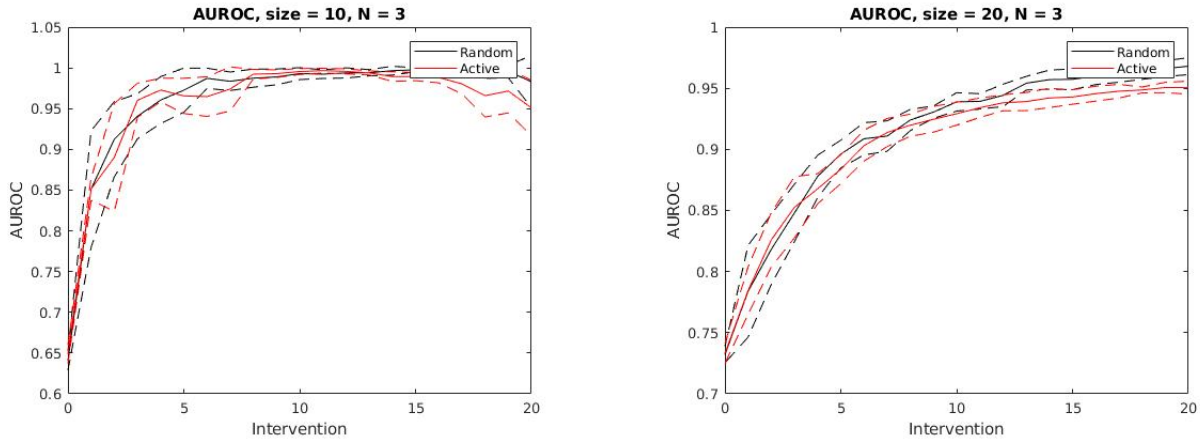


Figure 3. Replication study at size = 14. Note that active vs random are essentially tied in recovering network information.

Experiment 2: Naive batched strategy

During this set of computational experiments, I implemented the naive batched strategy as mentioned during the Background and Methods section. To recap, instead of simply taking the top most informative intervention, as predicted using the KL divergence, I performed the top N most informative interventions in parallel (calling this a batch). The results of these experiments at $N = 3$ can be seen in Figure 4. Judging on the results from the previous set of experiments, the right-hand plot in Figure 4 is not surprising, as we expect that this method will not perform well on nets larger than size = 14. However,

the left-hand plot is somewhat surprising. Here, we can see that the random and active strategies are essentially tied, with the active strategy even decreasing in AUROC by the last few interventions. This points to potential correlation among the ranked interventions, i.e., that the ranked experiments may be maximally informative, but that they are maximally informative *about the same thing*. Thus, after using the top experiment, we get no further information gain by using the subsequent ranked experiments, thus leading to worse performance than random. After seeing this phenomenon, I designed Experiment 3 to attempt to counteract correlation.



(a) Size = 10 **(b)** Size = 20

Figure 4. Naive batched strategy performed with 5 trials each, attempting to recover the same network during each trial.

Experiment 3: Attempting to counteract correlation

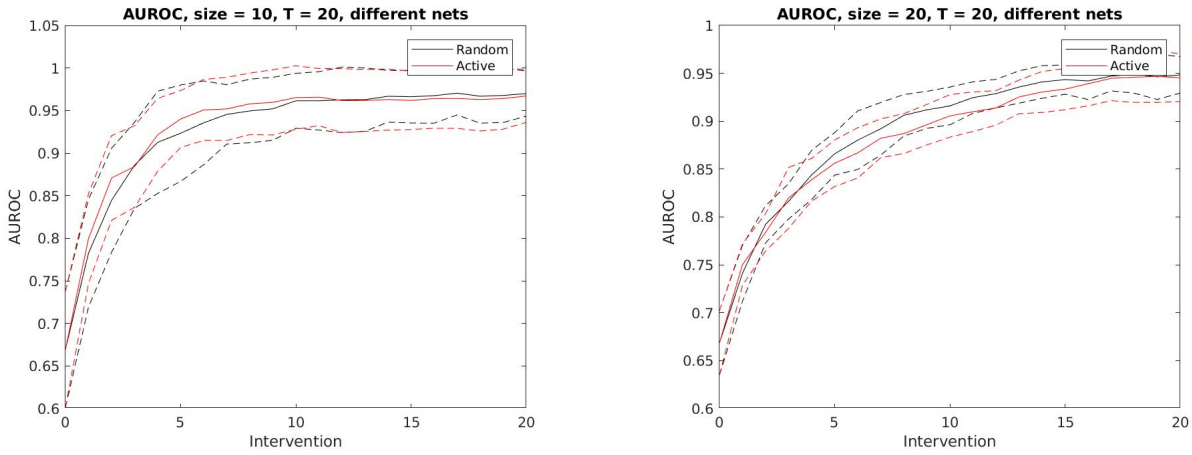
At this stage of experiments, I attempted to remove the 'correlation effect' observed in the active strategy of Experiment 2 by still utilizing the most informative intervention as predicted, but then using random interventions for the rest of the batch if $N > 1$. The random strategy was the same as in Experiment 2, utilizing random interventions for the entire batch. Figure 5 demonstrates that the effect of the actively selected intervention over all random interventions is essentially negligible - both plots in Figure 5 show essentially overlapping recovery. It's hard to determine whether correlation was present in Experiment 2 based on these results, as the effect of the actively selected intervention is lost among the random interventions. As can be seen, I only did experiments at $N = 3$ - it's possible that I could have seen a greater contrast between active and random (in terms of correlation effect) for smaller N, where the active intervention is less likely to be drowned out by the random interventions.

Experiment 4: Effects of increasing noise

In all of the above experiments, I used a constant noise variance of $\sigma^2 = 0.0025$. As I had not been able to demonstrate the efficacy of my naive batched strategy in previous experiments, I decided instead to investigate the effects of increased noise on the active learning strategy, using $\sigma^2 = 2.5$. As can be seen in Figure 6, the active learning strategy actually seems to *improve* with increased noise in measurements. These figures can be compared and contrasted with Figures 2 and 3, as these are the exact same experimental conditions, only with the increased noise. As opposed to the significant advantage that random had over active with low noise at size = 20, we find that active and random are essentially tied in recovery ability, while at size = 14, where originally random and active were tied, we now find that active seems to have a distinct advantage. This phenomenon begs for future investigation.

Discussion

In my original project proposal, I outlined three different Aims to investigate the efficacy of batched OED strategy. In particular, I wanted to look at how a constant batch size N changed recovery, how a batch size that increased or decreased over time changed recovery, or how an adaptive batch size changed recovery. Furthermore, I had planned to run computational experiments not only on randomly generated Bayesian networks, but also networks as generated by GeneNetWeaver[9] (GNW), in order to test networks that might be somewhat homologous to biological networks. During the course of my investigation, I only made it to the first Aim, and only tested on randomly generated networks. I unfortunately was never able to install GNW on my home computer and so decided to only focus on the recovery of randomly generated networks. Furthermore, many GNW-generated



(a) Size = 10

(b) Size = 20

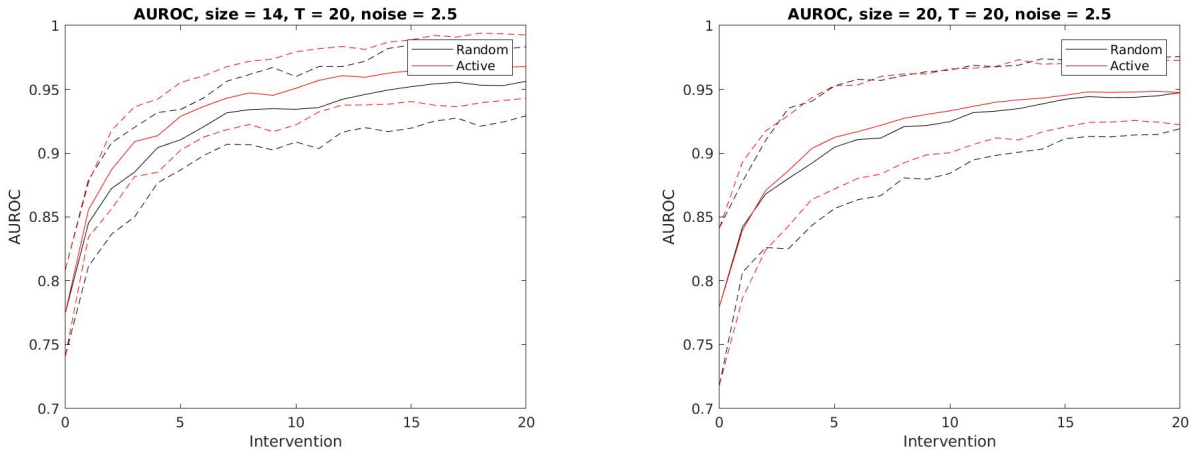
Figure 5. Attempts to remove correlation performed with 10 trials each, attempting to recover a different network during each trial.

networks contain loops and would have had to be thrown out anyway. Another barrier that kept me from achieving all of my project Aims was computational time: recovery of larger nets (size close to 20) could take several hours for several trials. Many of the experiments would thus be run overnight; however, it was quite difficult to sit down and make adjustments in real time to my project methods and approaches.

One of the most interesting phenomena observed was that of the original active learning framework failing for networks larger than size = 14. This was unforeseen and requires more theoretical analysis of the assumptions on which the original active learning method is based. However, in correspondence with my advisor Dr. Niall Mangan, we discussed the possibility that this method may rely on sampling techniques that are more viable in lower-dimensional spaces, but sample less well in higher-dimensional spaces. This is a hypothesis that requires further investigation in order to make this active learning strategy actually viable for the recovery of biological networks. Furthermore, as observed in Experiment 4, increased noise actually increases the ability of this strategy - again, this is a topic that requires further investigation.

Another phenomenon encountered was the correlation observed between experiments. While not completely unexpected (this strategy was ‘naive’, after all), this correlation had a stronger effect than I would have anticipated. Attempts to counteract this correlation resulted in random interventions ‘drowning out’ the effect of the active intervention, resulting in no advantage by active selection, as seen in Experiment 3.

Currently, it seems that naive parallelization of interventions is not a viable procedural policy for the recovery of biological Bayesian networks. However, there are still open questions surrounding these methods. Beyond the questions concerning active learning for networks of different size, does a parallelization strategy exist that only relies on the induced ranking by optimization strategies? Would ‘thinning’ the experiments (analogous to thinning in MCMC methods) help reduce correlation? Are there methods for predicting the correlation of interventions, using the current model structure? Why does noise affect active recovery so significantly? Could this noise mechanism be used to decrease correlation between interventions? All these and more are questions worthy of future investigation.



(a) Size = 14

(b) Size = 20

Figure 6. Increasing noise, performed with 10 trials each, attempting to recover a different network during each trial.

References

1. Cho, H., Berger, B. & Peng, J. Reconstructing causal biological networks through active learning. *PLoS ONE* **11** (2016).

2. Murphy, K. P. Active learning of causal bayes net structure (2001). Technical Report.

3. He, Y.-b. & He Yang-Bo, G. Z. Active learning of causal networks with intervention experiments and optimal designs. *J. Mach. Learn. Res.* **9**, 2523–2547 (2008).

4. Pournara, I. & Wernisch, L. Reconstruction of gene networks using Bayesian learning and manipulation experiments. *Bioinforma. (Oxford, England)* **20**, 2934–2942 (2004).

5. Ideker, T. E., Thorsson, V. & Karp, R. M. Discovery of regulatory interactions through perturbation: inference and experimental design. In *Pacific Symposium on Biocomputing 5*, 305–316 (2000).

6. Hauser, A. & Peter, B. Two optimal strategies for active learning of causal models from interventions. *Work. on Probabilistic Graph. Model.* 123–130 (2012).

7. Sverchkov, Y. & Craven, M. A review of active learning approaches to experimental design for uncovering biological networks. *PLoS Comput. Biol.* **13** (2017).

8. Szczurek, E., Gat-Viks, I., Tiuryn, J. & Vingron, M. Elucidating regulatory mechanisms downstream of a signaling pathway using informative experiments. *Mol. Syst. Biol.* **5** (2009).

9. Schaffter, T. Genenetweaver.

Acknowledgements

The author would like to acknowledge Niall Mangan for a helpful discussion on sampling in high-dimensional probability spaces.