A COMPARING COMPLEXITY CONSTANTS

A.1 About the classic instance for linear Top-m identification

In higher dimensions, and when $m = K - 2$, ω can be seen as the angle between the $m+1$-best arm vector and the hyperplane formed by the m best arm feature vectors. In order to check if, for $m \geq 1$, decreasing the value of $\omega \in (0, \frac{\pi}{2})$ yields to harder instances (as it is for $m = 1$), we ran the bandit algorithms on the instance $K = 4$, $N = 3$, $m = 2$ for $\omega \in \{\frac{\pi}{3}, \frac{\pi}{6}\}$. The resulting boxplots are shown in Figure 2. It can then be seen that indeed, for all algorithms, the empirical average sample complexity increases as ω decreases, which is an argument in favour of the use of this type of instance for the test of linear Top-m algorithms.

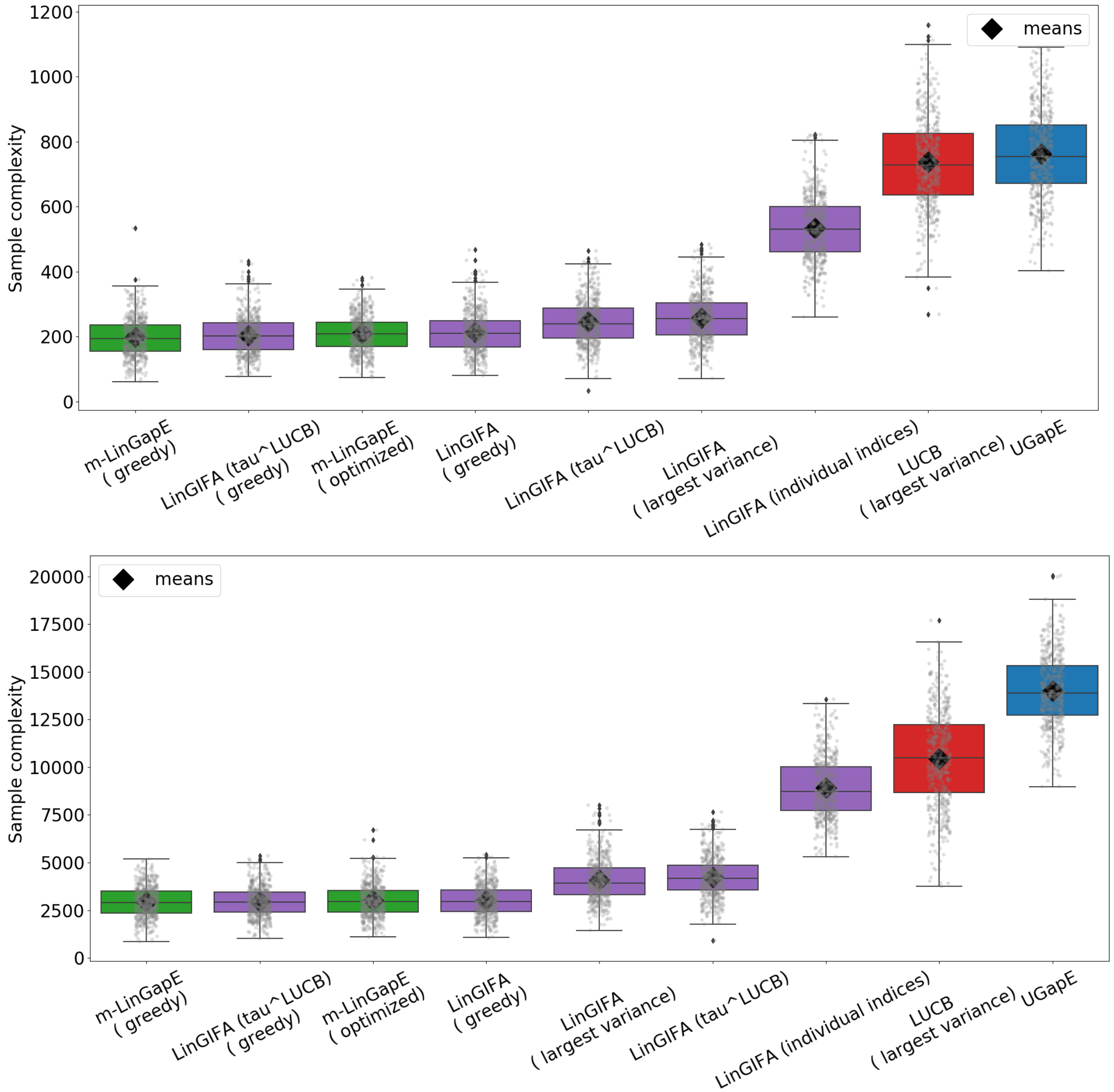

Figure 2: From top to bottom: classic instances (a) $K = 4$, $\omega = \frac{\pi}{3}$, $m = 2$; (b) $K = 4$, $\omega = \frac{\pi}{6}$, $m = 2$. Error frequencies are rounded up to 5 decimal places.

A.2 Comparing complexity constants

Below is the table to which we refer to in Section 4.4.

Table 4: Comparison of complexity constants in m-LinGapE and UGapE (% on $1,000$ random instances).

D	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.25	0.5	0.25	0.5	0.25	0.5	0.25	0.5
K	10	10	10	10	10	10	20	20	20	20	20	30	30	30	30
m	4	4	4	4	4	4	7	7	7	7	7	11	11	11	11
N	5	5	10	10	20	20	10	20	20	40	40	15	15	30	30
%	29.1%	30.8%	0.0%	0.0%	0.0%	0.0%	0.6%	0.0%	0.0%	0.0%	0.0%	0.1%	0.1%	0.0%	0.0%

We have tested if, empirically, LinGIFA was more performant than LinGapE on instances where $\mathrm{H}^{\varepsilon}(m\text{-LinGapE}(2), \mu) \leq \mathrm{H}^{\varepsilon}(\mathrm{UGapE}, \mu)$, since LinGIFA has a similar structure as UGapE. We generated a random linear instance, following the procedure described in Section 4.4 in the paper, with $K = 10$, $N = 5$, $D = 0.5$. For $m = 3$, the condition $\mathrm{H}^{\varepsilon}(m\text{-LinGapE}(2), \mu) \leq \mathrm{H}^{\varepsilon}(\mathrm{UGapE}, \mu)$ is satisfied, whereas it is not when $m = 8$. We considered Gaussian reward distributions. See Figure 3. From these results, we notice that both algorithms are actually similar in sample complexity in both instances. Hence, even if the condition $\mathrm{H}^{\varepsilon}(m\text{-LinGapE}(2), \mu) \leq \mathrm{H}^{\varepsilon}(\mathrm{UGapE}, \mu)$ is seldom satisfied as seen in Table 4, in practice, m-LinGapE with the optimized rule is still performant.

Table 5: Values of complexity constants in m-LinGapE and UGapE on the randomly generated instance.

	$m = 3$	$m = 8$
$\mathrm{H}^{\varepsilon}(m\text{-LinGapE}(2), \mu)$	$4,545.97$	$32,124.01$
$\mathrm{H}^{\varepsilon}(\mathrm{UGapE}, \mu)$	$5,047.76$	$27,622.18$
$\mu_m - \mu_{m+1}$	0.075	0.029
$\mathrm{H}^{\varepsilon}(m\text{-LinGapE}(2), \mu) \leq \mathrm{H}^{\varepsilon}(\mathrm{UGapE}, \mu)$?	**True**	**False**

B DRUG REPURPOSING INSTANCE

Remember that we call "phenotypes" gene activity profiles of patients and controls (healthy group) (that is, vectors which represent the genewise activity in a finite set of genes). We focus on a finite set of genes called M30, which has been shown to have a global gene activity that is anti-correlated to epileptic gene activity profiles (Delahaye-Duriez et al., 2016). The bandit instance comprises of arms/drugs, which, once pulled, return a single score/reward which quantifies their ability to "reverse" a epileptic phenotype –that is, an anti-epileptic treated epileptic phenotype should be closer to a healthy phenotype. A gene regulatory network (GRN) is a summary of gene transcriptomic interactions as a graph: nodes are genes or proteins, (directed) edges are regulatory interactions. We see a GRN as a Boolean network (BN): nodes can have two states (0 or 1), and each of them is assigned a so-called "regulatory function", that is, a logical formulæ which updates their state given the states of regulators (i.e., predecessors in the network) at each time step. In order to infer the effect of a treatment, one can set as initial network state ("initial condition") the patient phenotype masked by the perturbations on the drug targets, and iteratively update the network state until reaching an attractor state (`phenotype_prediction` procedure).

Building the Boolean network We use the Boolean network inference method described in (Yordanov et al., 2016) using code at repository `https://github.com/regulomics/expansion-network` (Réda and Wilczyński, 2020). We get the unsigned undirected regulatory interactions from the *protein-protein interaction network* (PPI) of M30, using the STRING database (Szklarczyk et al., 2016). Using expression (or, as we called it in the paper, gene activity) data in the hippocampus from UKBEC data (Gene Expression Omnibus (GEO) accession number $GSE46706$), a Pearson's R correlation matrix is computed, which allows signing the interactions using pairwise correlation signs. Then, considering that the effects of a gene perturbation can only be seen on connected nodes, we only keep strongly connected components in which at least one gene perturbation in LINCS L1000 experiments occurs, using Tarjan's algorithm (Tarjan, 1972).

Then, in order to direct the edges in the network using the inference method, we restrict the experiments extracted from LINCS L1000 to those in SH-SY5Y human neuron cells (neuroblastoma from bone marrow), with a positive interference scale score (Cheng and Li, 2016), which quantify the success of the perturbation experiment. For each experiment (knockdown via shRNA, overexpression via cDNA) on a gene in M30 in this cell line, we extract from Level 3 LINCS same-plate untreated, genetic control and perturbed profiles (each of them being real-valued vectors of size ≈ 100, the number of genes in M30) such as the perturbed profile on this plate has the largest value of *distil_ss* which quantifies experimental replication. This procedure yields a total of (1 untreated + 2 replicates of genetic control + 2 replicates of perturbed) ×3 experimental profiles. Each experimental constraint for the GRN inference is defined as follows ($G = 101$ is the number of $M30$ genes in the network):

- **Initial condition**: Untreated profile which has been binarized using the `binarize_via_histogram` procedure (peak detection in histogram of gene expression values using persistent topology). Value: $\{0, 1, \perp\}^G$. The number of non-$\perp$ values is 66.
- **Perturbation**: Gene-associated value is equal to 1 if and only if the gene is perturbed in the experiment.
- **Final/fixpoint condition**: Vector which has been obtained by running Characteristic Direction (Clark et al., 2014) (CD) on [treated||genetic control] (in the call to the function, treated profiles were annotated 2 whereas genetic control ones were annotated 1) profiles, which yields a vector in $\{0, 1, \perp\}^G$, where 0's (resp. 1's) are significantly down- (resp. up-) regulated genes in treated profiles compared to control ones. Note that $[P_1||P_2]$ means that we compute the genewise activity change from group P_2 to group P_1 (hence, this is not symmetrical). The number of non-$\perp$ values is around 22.

The inferred GRN should satisfy all experimental constraints by assigning logical functions to genes and selecting gene interactions. This network is displayed in Figure 4.

Table 6: Experiments in SH-SY5Y human neuron-like cell line for GRN inference. Inference parameter: $t = 20$, asynchronous dynamics.

Perturbed gene	**Experiment type**	**Exposure time**
CACNA1C	KD	120 h
CDC42	KD	120 h
KCNA2	KD	120 h

Getting the arm/drug features The features we use are the drug signatures ($K = 509$ in the binary drug signature dataset). Given a drug, we compute them as follows:

1. First, in the Level 3 LINCS L1000 database (Subramanian et al., 2017), we select the cell line with the highest transcriptomic activity score, or TAS (quantifying the success of treatment in this specific cell line: we expect to obtain more reproducible experiments in this cell line if this score is high). Then, considering experiment in this cell line, treatment by the considered drug, we select the *brew_prefix* (identifier for experimental plate) which correspond to the treated expression profile with the largest value of *distil_ss* (which quantifies the reproducibility of the profile across replicates), and we get the corresponding same-plate control (vehicle) profile. We also get one same-plate replicate of the considered treated experiment and another of the control experiment (total number of profiles: $2 + 2$).
2. We apply on this set of profiles Characteristic Direction (Clark et al., 2014) in order to get the relative genewise expression change due to treatment from control sample group CD([treated||vehicle control]). This yields a real-valued vector in $[-1, 1]^G$ which will be used in the baseline method L1000 CDS^2, and a binary vector in $\{0, 1, \perp\}^G$ in our scoring method, which is the so-called drug signature.

Epileptic patient/control phenotypes We fetched data from GEO accession number $GSE77578$, which was then quantile-normalized across all patient ($|P_p| = 18$) and control ($|P_c| = 17$) samples. We run Characteristic Direction (Clark et al., 2014) $CD([P_c||P_p])$ in order to get the "differential phenotype" from controls to patients, which is the way we chose in order to aggregate control profiles and only considering differentially expressed genes.

Drug "true" scores We get them from RepoDB (Brown and Patel, 2017) database, which is a curated version from `clinicaltrials.gov`, and from literature. To each drug is associated an integer: 1 if the drug is antiepileptic, 0 if it is unknown, -1 if it is a proconvulsant drug.

Masking procedure $\rtimes$ We use this (asymmetric) function in order to generate the initial condition from which an attractor state, if it exists, should be fetched: $(x \rtimes y)[j] = y[j]$ if $y[j] \in \{0, 1\}$ else $x[j]$. This aims at mimicking the immediate effect of treatment on gene activity.

Running the simulator *via* the GRN Given collected patient and control phenotypes, and seeing arms as potentially repurposed drugs, the procedure to generate a reward from a given arm a is as described in Algorithm 2. We compare this method to a simpler signature reversion method, used in the web application L1000 CDS^2 (Duan et al., 2016), which is deterministic, and compares directly drug signatures and differential phenotypes. The full procedure is described in Algorithm 3. We have tested our method on a subset of drugs with respect to this baseline. The results can be seen in Figure 5.

Note that returning a score for a single drug with our method is usually a matter of a few minutes, but the computation time can drastically increase when considering a higher number of nodes in the Boolean network, so that is why, even if on this instance we could run all computations for each drug and for each initial patient sample for drug repurposing (which is what we do in Figure 5 anyway in order to check that our method yields correct results with respect to known therapeutic indications), we think this model is interesting to test our bandit algorithms.

Moreover, the linear dependency between features and scores does not hold: indeed, in our subset of $K = 10$ arms, computing the least squares estimate of θ using the mean rewards $\overline{m}$ as true values, and denoting X the concatenation of drug signatures, that is, $\theta = (X^\top X)^{-1} X \overline{m}$, gives a high value of $\|\theta - \overline{m}\| \approx 20.9$. The linear setting in bandits is simply the easiest contextual setting to analyze. Although this non-linearity, along with the fact that the initial condition is randomized, might be the main reason why the empirical sample complexity in our subset instance is a lot higher than 10×18 for all algorithms, even if linear algorithms are noticeably more performant than classical ones.

Algorithm 2 Reward generation via the “GRN simulator”.

```
requires $G$ GRN, phenotypes of patient (diseased) and control (healthy) individuals w.r.t. a given disease $P_p \in [0, 15]^G$, $P_c \in [0, 15]^G$, $a$ arm/drug to be tested, with binary drug signature $s^b_a \in \{0, 1, \perp\}^G$.
# differential phenotype is computed: controls||patients
$D \in \{0, 1, \perp\}^G = CD([P_c||P_p])$
# patient phenotype is uniformly sampled from the pool of patient phenotypes
$p \sim \mathcal{U}(P_p)$
$p^b \in \{0, 1, \perp\}^G \leftarrow$ binarize_via_histogram($p$)
$p^r \in \{0, 1\}^G \leftarrow$ phenotype_prediction(GRN=G, initial_condition=$(p^b \rtimes s^b_a) \in \{0, 1, \perp\}^G$)
# comparison function cosine_score is run on the intersection of supports of $D$ and $p^r$
# this intersection is equal to $50$ in practice, which is the size of the support of $D$
$r \leftarrow$ cosine_score$_{|D|\cap|p^r|}(D, p^r)$
returns $r$
```

Algorithm 3 Reward via baseline method from L1000 CDS^2 (Duan et al., 2016).

```
requires Phenotypes of patient (diseased) and control (healthy) individuals w.r.t. a given disease $P_p \in [0, 15]^G$, $P_c \in [0, 15]^G$, $a$ arm/drug to be tested, with non-binary, full signature $s_a \in [-1, 1]^G$.
# differential phenotype is computed: patients||controls
$C \in \{0, 1, \perp\}^G = CD([P_p||P_c])$
$r \leftarrow 1 -$ cosine_score$(C, s_a)$
returns $r$
```

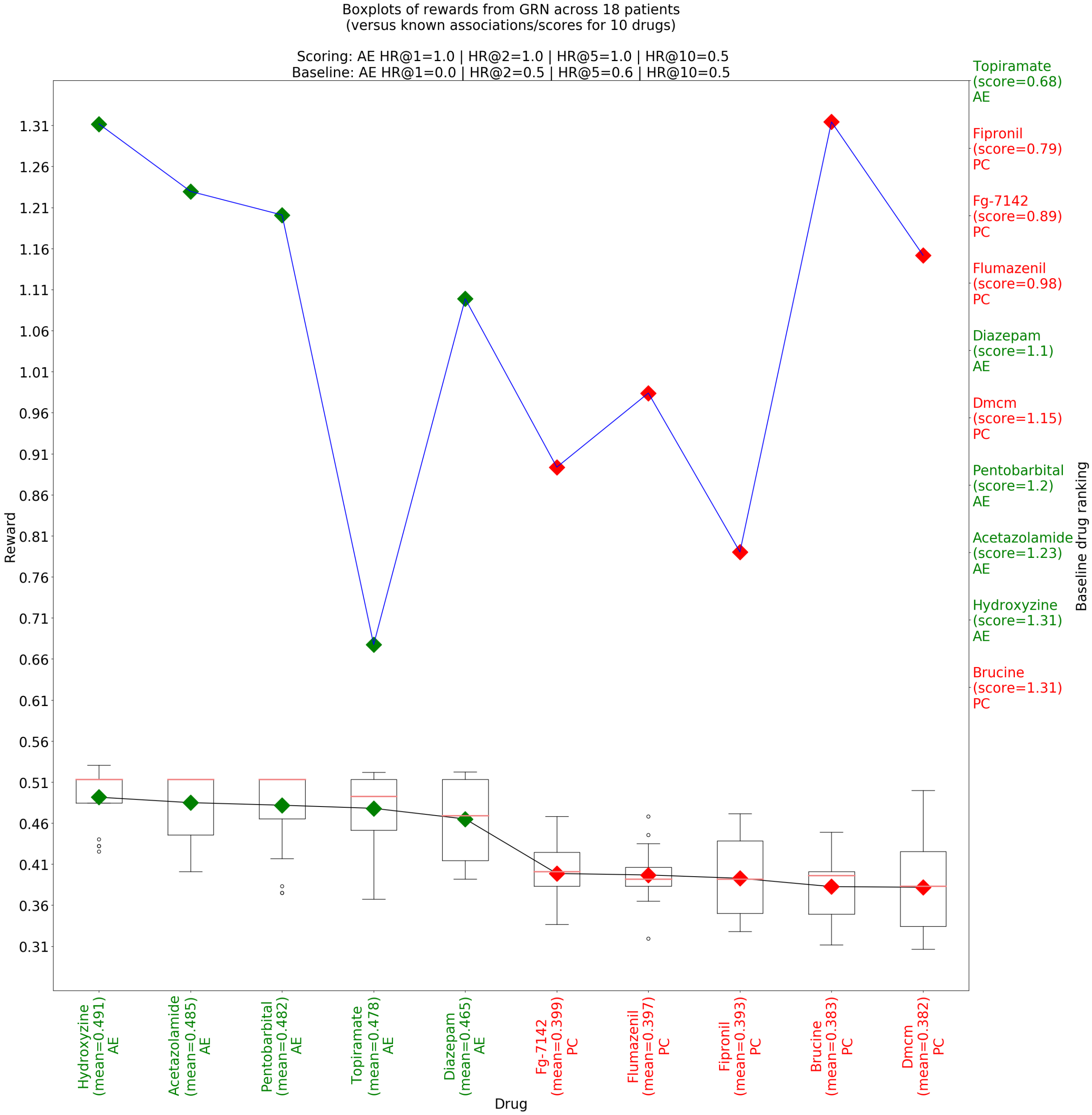

Figure 5: We consider a subset of drugs of size 10 (5 with positive association score, 5 with negative score), which is the one tested in the paper. For validation, we plot a boxplot of the rewards obtained for each initial patient sample, for each drug. Mean is colored as green if the drug is antiepileptic (AE), resp. red if it is proconvulsant (PC), with the corresponding drug name (in red if its true score is negative, in green if it is positive). The baseline score is plot in in blue. For both methods, the highest the score is, the better (the “more” anti-epileptic the drug is predicted). We computed and reported above the plot the Hit Score at rank r (HR@r), that is, the mean accuracy on the class AE on the Top-r scores, for $r \in \{1, 2, 5, 10\}$ for each of the methods (Scoring or Baseline).

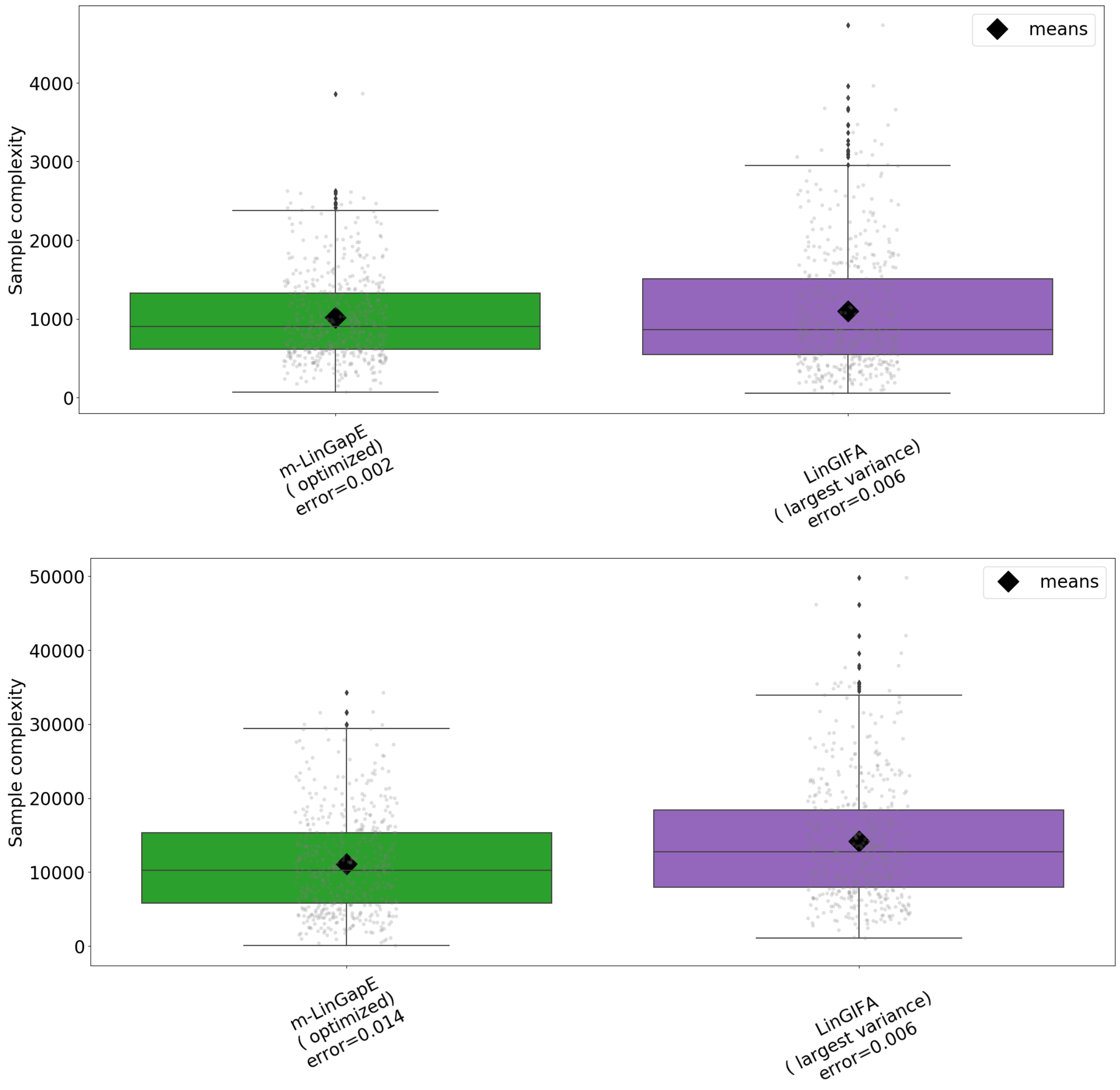

Figure 3: From top to bottom: $m = 3$, $m = 8$. Error frequencies are rounded up to 5 decimal places.

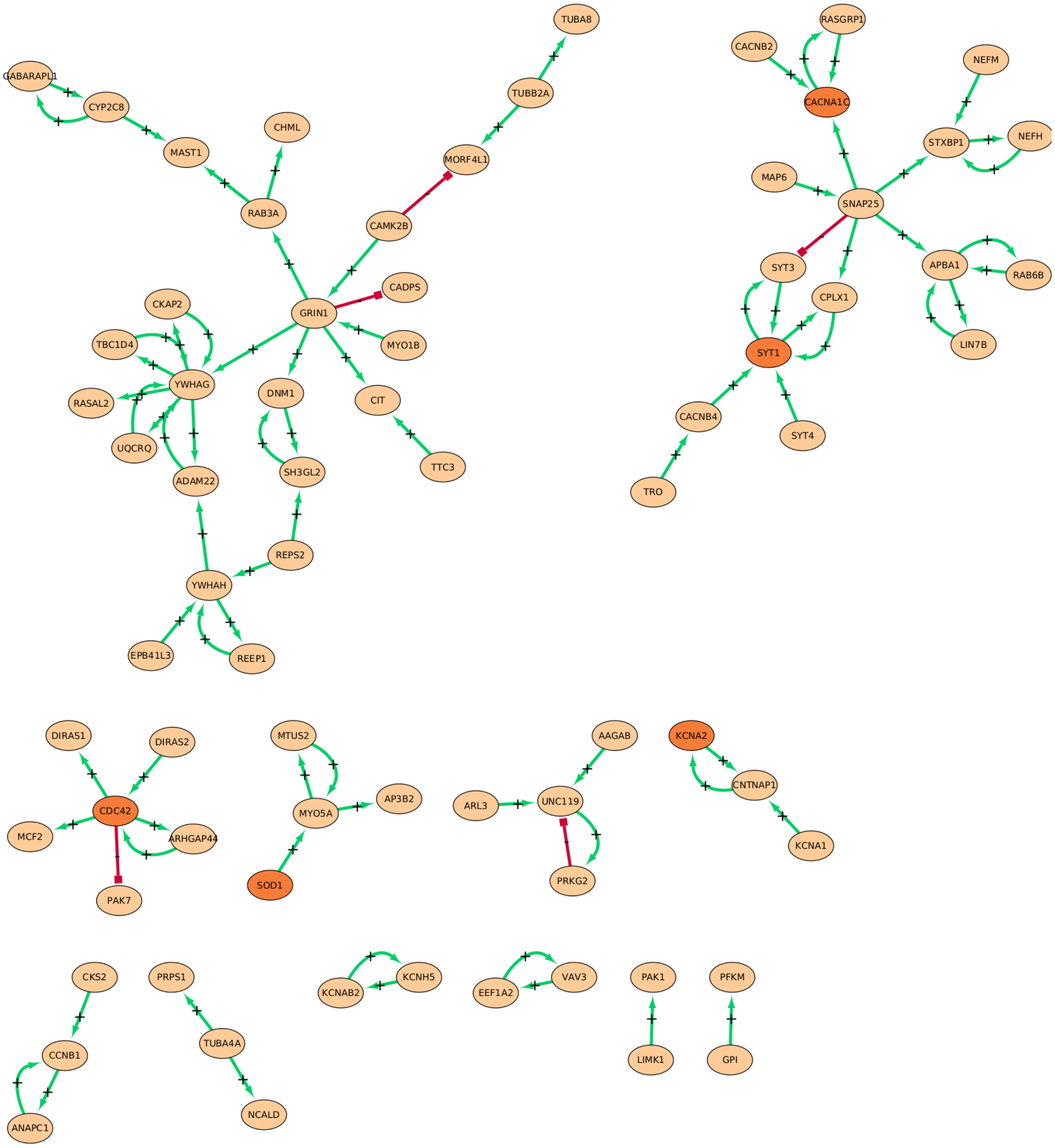

Figure 4: Inferred GRN for the drug repurposing instance on epilepsy. Genes present in this network belong to the M30 set. Green edges labelled “+” (resp. red edges labelled “-”) are activatory (resp. inhibitory) interactions from one regulatory gene on a target gene, that is, that increase (resp. decrease) target gene activity. Deep orange nodes are perturbed nodes in the SHSY5Y cell line in LINCS L1000.

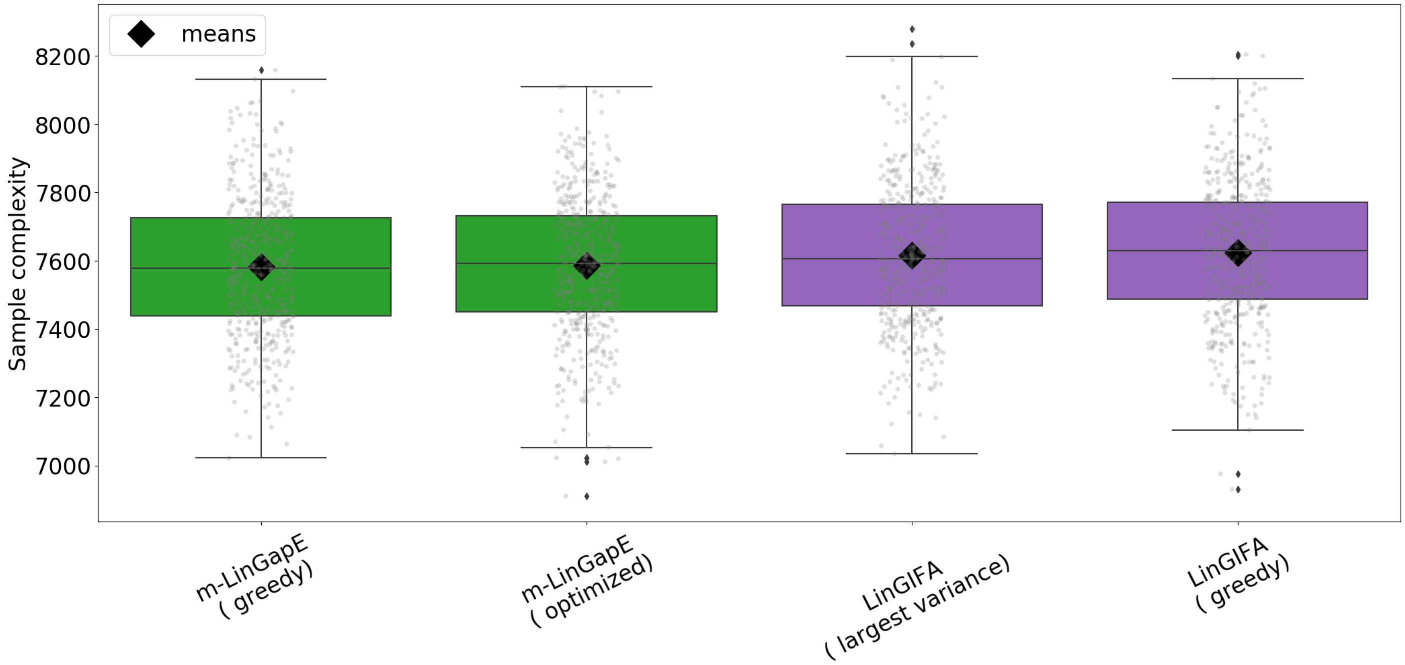

Figure 6: Drug repurposing instance $K = 10$, 500 simulations, $m = 5$, $\delta = 0.05, \epsilon = 0, \sigma = 0.5, \lambda = \sigma/20$. Close-up from Figure 1 from the paper.

C UPPER BOUNDS FOR GIFA ALGORITHMS

Lemma 7. *In algorithm m-LinGapE, for any selection rule, on event* $\mathcal{E} \triangleq \bigcap_{t>0} \bigcap_{i,j\in[K]} \Big(\Delta_{i,j} \in [-B_{j,i}(t), B_{i,j}(t)]\Big)$*, for all* $t > 0$*,* $B_{c_t,b_t}(t) \leq \min(-(\Delta_{b_t} \vee \Delta_{c_t}) + 2W_t(b_t, c_t), 0) + W_t(b_t, c_t)$*.* (Lemma 4 in the paper)

Proof. Let us use two properties:

1. As $b_t \in J(t)$ and $c_t \notin J(t)$, it holds in particular that $\hat{\mu}_{b_t}(t) \geq \hat{\mu}_{c_t}(t)$, hence $B_{c_t,b_t}(t) = \hat{\Delta}_{c_t,b_t}(t) + W_t(b_t, c_t) \leq W_t(b_t, c_t)$.

2. From the definitions of b_t and c_t, it holds that $B_{c_t,b_t}(t) = \max_{j\in J(t)} \max_{i\notin J(t)} B_{i,j}(t)$.

Property 1 already establishes that $B_{c_t,b_t}(t) \leq W_t(b_t, c_t)$, it therefore remains to show that $B_{c_t,b_t}(t) \leq -(\Delta_{b_t} \vee \Delta_{c_t}) + 3W_t(b_t, c_t)$. We do it by distinguishing four cases:

(i) $b_t \in \mathcal{S}^\star_m$ **and** $c_t \notin \mathcal{S}^\star_m$: In that case $\Delta_{b_t} = \mu_{b_t} - \mu_{m+1}$ and $\Delta_{c_t} = \mu_m - \mu_{c_t}$. As event $\mathcal{E}$ holds, one has $B_{c_t,b_t}(t) = -B_{b_t,c_t}(t) + 2W_t(b_t, c_t) \leq \Delta_{c_t,b_t} + 2W_t(b_t, c_t)$. As $c_t \notin \mathcal{S}^\star_m$, $\mu_{c_t} \leq \mu_{m+1}$, and $\Delta_{c_t,b_t} \leq \mu_{m+1} - \mu_{b_t} = -\Delta_{b_t}$. But as $b_t \in \mathcal{S}^\star_m$, it also holds that $\mu_{b_t} \geq \mu_m$, and $\Delta_{c_t,b_t} \leq \mu_{c_t} - \mu_m = -\Delta_{c_t}$. Hence $B_{c_t,b_t}(t) \leq -(\Delta_{b_t} \vee \Delta_{c_t}) + 2W_t(b_t, c_t) \leq -(\Delta_{b_t} \vee \Delta_{c_t}) + 3W_t(b_t, c_t)$.

(ii) $b_t \notin \mathcal{S}^\star_m$ **and** $c_t \in \mathcal{S}^\star_m$: Using Property 1:

$$B_{c_t,b_t}(t) \leq W_t(b_t, c_t) \leq \hat{\Delta}_{b_t,c_t}(t) + W_t(b_t, c_t) = B_{b_t,c_t}(t) = -B_{c_t,b_t}(t) + 2W_t(b_t, c_t) \leq \Delta_{b_t,c_t} + 2W_t(b_t, c_t)$$

as event $\mathcal{E}$ holds. One can show with the same arguments as in the previous case that $B_{c_t,b_t}(t) \leq -(\Delta_{b_t} \vee \Delta_{c_t}) + 3W_t(b_t, c_t)$.

(iii) $b_t \notin \mathcal{S}^\star_m$ **and** $c_t \notin \mathcal{S}^\star_m$: In that case, there must exist $b \in \mathcal{S}^\star_m$ that belongs to $J(t)^c$. From the definition of c_t, it follows that $B_{c_t,b_t}(t) \geq B_{b,b_t}(t)$. Hence, using furthermore Property 1, the definition of c_t, event $\mathcal{E}$ and $b \in \mathcal{S}^\star_m$, $W_t(b_t, c_t) \geq B_{c_t,b_t}(t) \geq B_{b,b_t}(t) \geq \Delta_{b,b_t} \geq \Delta_{m,b_t} = \Delta_{b_t}$. It follows that, using event $\mathcal{E}$:

$$\begin{aligned} B_{c_t,b_t}(t) &\leq \Delta_{c_t,b_t} + 2W_t(b_t, c_t) = (\mu_{c_t} - \mu_m) + (\mu_m - \mu_{b_t}) + 2W_t(b_t, c_t) = -\Delta_{c_t} + \Delta_{b_t} + 2W_t(b_t, c_t) \\ &\quad (b_t \notin \mathcal{S}^\star_m \text{ and } c_t \notin \mathcal{S}^\star_m) \\ &\leq -\Delta_{c_t} + 3W_t(b_t, c_t) \end{aligned}$$

And it also holds by Property 1 that:

$$\begin{aligned} B_{c_t,b_t}(t) &\leq W_t(b_t, c_t) = -W_t(b_t, c_t) + 2W_t(b_t, c_t) \\ &\leq -\Delta_{b_t} + 2W_t(b_t, c_t) \leq -\Delta_{b_t} + 3W_t(b_t, c_t) \end{aligned}$$

Hence $B_{c_t,b_t}(t) \leq -(\Delta_{b_t} \vee \Delta_{c_t}) + 3W_t(b_t, c_t)$.

(iv) $b_t \in \mathcal{S}^\star_m$ **and** $c_t \in \mathcal{S}^\star_m$: In that case, there must exist $c \notin \mathcal{S}^\star_m$ such that $c \in J(t)$. By Property 2, on event $\mathcal{E}$ and using $c \in (\mathcal{S}^\star_m)^c$, we know that $B_{c_t,b_t}(t) = \max_{j\in J(t)} \max_{i\notin J(t)} B_{i,j}(t) \geq \max_{i\notin J(t)} B_{i,c}(t) \geq B_{c_t,c}(t) \geq \mu_{c_t} - \mu_c \geq \mu_{c_t} - \mu_{m+1} = \Delta_{c_t}$. Hence, using furthermore Property 1 yields $\Delta_{c_t} \leq B_{c_t,b_t}(t) \leq W_t(b_t, c_t)$. It follows that, using event $\mathcal{E}$:

$$\begin{aligned} B_{c_t,b_t}(t) &\leq \mu_{c_t} - \mu_{b_t} + 2W_t(b_t, c_t) = \mu_{c_t} - \mu_{m+1} + \mu_{m+1} - \mu_{b_t} + 2W_t(b_t, c_t) \\ &= \Delta_{c_t} - \Delta_{b_t} + 2W_t(b_t, c_t) \leq -\Delta_{b_t} + 3W_t(b_t, c_t) \end{aligned}$$

And using again Property 1, one has:

$$\begin{aligned} B_{c_t,b_t}(t) &\leq W_t(b_t,c_t) = -W_t(b_t,c_t) + 2W_t(b_t,c_t) \leq -\Delta_{c_t} + 2W_t(b_t,c_t) \\ &\leq -\Delta_{c_t} + 3W_t(b_t,c_t) \end{aligned}$$

Hence $B_{c_t,b_t}(t) \leq -(\Delta_{b_t} \vee \Delta_{c_t}) + 3W_t(b_t,c_t)$, which is what we wanted to show. □

Lemma 8. Upper bound in m-LinGapE with either or both b_t and c_t pulled at time t (m-LinGapE(1)) *Maximum number of samplings on event $\mathcal{E}$ is upper-bound by* $\inf_{u \in \mathbb{R}^{*+}}\{u > 1 + H^{\varepsilon}(m\text{-LinGapE(1)}, \mu)C^2_{\delta,u}\}$*, where* $H^{\varepsilon}(m\text{-LinGapE(1)}, \mu) \triangleq 4\sigma^2 \sum_{a \in [K]} \max\left(\varepsilon, \frac{\varepsilon+\Delta_a}{3}\right)^{-2}$.

Proof. Combining Lemma 4 with stopping rule τ^{LUCB}, at time $t < \tau^{LUCB}$:

$$\begin{aligned} \varepsilon \leq B_{c_t,b_t}(t) &\leq \min(-(\Delta_{b_t} \vee \Delta_{c_t}) + 3W_t(b_t,c_t), W_t(b_t,c_t)) \\ \Leftrightarrow \max\left(\varepsilon, \frac{\varepsilon+\Delta_{b_t}}{3}, \frac{\varepsilon+\Delta_{c_t}}{3}\right) &\leq W_t(b_t,c_t) \leq W_t(b_t) + W_t(c_t) \leq 2W_t(a_t) = 2C_{\delta,t}||x_{a_t}||_{\hat{\Sigma}_t^{\lambda}} \\ &\text{(where } a_t = \max_{a \in \{b_t,c_t\}} W_t(a)) \\ &= 2\sigma C_{\delta,t}||x_{a_t}||_{(\hat{V}_t^{\lambda})^{-1}} \leq 2\sigma C_{\delta,t}\frac{||x_{a_t}||}{\sqrt{N_{a_t}(t)}||x_{a_t}||} \\ &= 2\sigma C_{\delta,t}\frac{1}{\sqrt{N_{a_t}(t)}} \text{ (using Lemma 2 and } \lambda > 0, N_{a_t}(t) > 0\text{, since } a_t \text{ is pulled at } t) \\ \Leftrightarrow N_{a_t}(t) &\leq \frac{4\sigma^2 C^2_{\delta,t}}{\max\left(\varepsilon, \frac{\varepsilon+\Delta_{b_t}}{3}, \frac{\varepsilon+\Delta_{c_t}}{3}\right)^2} \leq \min_{a \in \{b_t,c_t\}} \frac{4\sigma^2 C^2_{\delta,t}}{\max\left(\varepsilon, \frac{\varepsilon+\Delta_a}{3}\right)^2} \leq \frac{4\sigma^2 C^2_{\delta,t}}{\max\left(\varepsilon, \frac{\varepsilon+\Delta_{a_t}}{3}\right)^2} \\ \Leftrightarrow N_{a_t}(t) &\leq \frac{4\sigma^2 C^2_{\delta,t}}{\max\left(\varepsilon, \frac{\varepsilon+\Delta_{a_t}}{3}\right)^2} = T^*(a_t,\delta,t) \end{aligned}$$

Using Lemma 6, if $T(\mu,\delta)$ is the number of samplings of m-LinGapE on bandit instance μ for δ-fixed confidence Top-m identification:

$$T(\mu,\delta) \leq \inf_{u \in \mathbb{R}^{*+}} \left\{ u > 1 + C^2_{\delta,u} \sum_{a \in [K]} \frac{4\sigma^2}{\max\left(\varepsilon, \frac{\varepsilon+\Delta_a}{3}\right)^2} \right\} \leq \inf_{u \in \mathbb{R}^{*+}} \{u > 1 + C^2_{\delta,u}\mathrm{H}^{\varepsilon}(m\text{-LinGapE}(1),\mu)\}$$

□

D TECHNICAL LEMMAS

Lemma 9. *Let us fix $K > m > 0$, $t > 0$ and $i \in [K]$. Let us consider μ such that $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_K$, and a series of distinct values $(B_{j,i}(t))_{j \in [K]}$ such that $B_{j,i}(t) \geq \mu_j - \mu_i$ for any $j \in [K]$. Then* $\overset{m}{\max_{j \in [K]}} B_{j,i}(t) \geq \mu_m - \mu_i$.

Proof. Assume by appealing to the extremes that $\overset{m}{\max_{j \in [K]}} B_{j,i}(t) < \mu_m - \mu_i$. Then, using our assumption on $(B_{j,i}(t))_{j \in [K]}$ and $(\mu_j)_{j \in [K]}$, for any $j \in [m]$, $B_{j,i}(t) \geq \mu_j - \mu_i \geq \mu_m - \mu_i > \overset{m}{\max_{j \in [K]}} B_{j,i}(t)$, which means at least m distinct values of $(B_{j,i}(t))_{j \leq K}$ are strictly greater than $\overset{m}{\max_{j \in [K]}} B_{j,i}(t)$, which yields a contradiction. Thus $\overset{m}{\max_{j \in [K]}} B_{j,i}(t) \geq \mu_m - \mu_i$. Note that we can assume the condition on $(B_{i,j}(t))_{i,j \in [K]}$ being distinct is satisfied except for some degenerate cases where two arm features are equal and the observations made from both arms are exactly the same. □

Lemma 10. *For all $t > 0$, for any subset $J \subseteq [K]$ of size m, for all $j \in J$, $\overset{m}{\underset{i\neq j}{\max}}\, B_{i,j}(t) \leq \max_{i\notin J} B_{i,j}(t)$.* (Lemma 1 in the paper)

Proof. Indeed, $\overset{m}{\underset{i\neq j}{\max}}\, B_{i,j}(t) = \min_{S\subseteq[K],|S|=m-1} \max_{i\notin(S\cup\{j\})} B_{i,j}(t)$ (set S matching the outer bound is $\underset{i\neq j}{\overset{[m-1]}{\arg\max}}\, B_{i,j}(t)$, meaning that we consider then the maximum value over the set of $(B_{i,j}(t))_{i\in[K]}$ from which the $m-1$ largest values and $B_{j,j}(t)$ are removed). Then consider $S = J\setminus\{j\}$, which is included in $[K]$ and is of size $m-1$ ($j \in J$). Then $\overset{m}{\underset{i\neq j}{\max}}\, B_{i,j}(t) \leq \max_{i\notin S\cup\{j\}} B_{i,j}(t) = \max_{i\notin J} B_{i,j}(t)$. □

Lemma 11. *For any $t > 0$, for any $a \in [K]$ such that $N_a(t) > 0$, for all $x \in \mathbb{R}^N$, $||x||^2_{(\hat{V}_t^\lambda)^{-1}} \leq x^\top(\lambda I_N + N_a(t)x_a x_a^\top)^{-1}x$.*

Proof. Let us prove this lemma by induction on $K \geq 2$ (case $K = 1$ is trivial). Let $A_t(a) \triangleq N_a(t)x_a x_a^\top$, $A_t \triangleq \lambda I_N + A_t(a)$. For $[K] = \{a_1, a_2, \ldots, a_{K-1}, a\}$ and $K \geq 2$, let us denote $B_K^t \triangleq \sum_{i=1}^{K-1} A_t(a_i)$, such that $\hat{V}_t^\lambda = A_t + B_K^t$. We will prove a stronger claim, which is "for any $t \in \mathbb{N}^*$, for any $x \in \mathbb{R}^N$, and $K \geq 2$, A_t and $A_t + B_K^t$ are invertible and $||x||^2_{(A_t+B_K^t)^{-1}} < ||x||^2_{A_t^{-1}}$". Note that, for any K and t, since $\lambda > 0$, A_t is then a Gram matrix with linearly independent columns, thus is positive definite, and B_K^t is a Gram matrix, thus a non-negative definite matrix. Then $A_t + B_K^t$ and A_t are positive definite and invertible.

<u>If $K = 2$:</u> then let us assume that $[K] = \{a, a_1\}$:

$$
\begin{aligned}
||x||^2_{(A_t+B_2^t)^{-1}} &\triangleq x^\top(A_t + B_2^t)^{-1}x = x^\top(A_t + N_{a_1}(t)x_{a_1}x_{a_1}^\top)^{-1}x \text{ (using Sherman-Morrison formula)}\\
&= x^\top\Big(A_t^{-1} - \frac{A_t^{-1}N_{a_1}(t)x_{a_1}x_{a_1}^\top A_t^{-1}}{1 + N_{a_1}(t)||x_{a_1}||^2_{A_t^{-1}}}\Big)x = ||x||^2_{A_t^{-1}} - \frac{(A_t^{-1}x)^\top B_2^t(A_t^{-1}x)}{1 + N_{a_1}(t)||x_{a_1}||^2_{A_t^{-1}}} \leq ||x||^2_{A_t^{-1}} - 0
\end{aligned}
$$

using the fact that B_2^t is nonnegative definite and A_t, and then A_t^{-1}, are both symmetric.

<u>If $K > 2$:</u> using the induction, $A_t + B_{K-1}^t$ is invertible. Similarly to the previous step, using the Sherman-Morrison formula:

$$
\begin{aligned}
||x||^2_{(A_t+B_K^t)^{-1}} &= x^\top(A_t + B_{K-1}^t)^{-1}x - \frac{x^\top(A_t + B_{K-1}^t)^{-1}A_t(a_{K-1})(A_t + B_{K-1}^t)^{-1}x}{1 + N_{a_{K-1}}(t)||x_{a_{K-1}}||^2_{(A_t+B_{K-1}^t)^{-1}}}\\
&= x^\top(A_t + B_{K-1}^t)^{-1}x - \frac{((A_t + B_{K-1}^t)^{-1}x)^\top A_t(a_{K-1})((A_t + B_{K-1}^t)^{-1}x)}{1 + N_{a_{K-1}}(t)||x_{a_{K-1}}||^2_{(A_t+B_{K-1}^t)^{-1}}}\\
&\text{(same argument as previously, since } A_t(a_{K-1}) \text{ is a Gram matrix)}\\
&\leq x^\top(A_t + B_{K-1}^t)^{-1}x - 0 = ||x||^2_{(A_t+B_{K-1}^t)^{-1}}
\end{aligned}
$$

Then, using the induction, $||x||^2_{(\hat{V}_t^\lambda)^{-1}} = ||x||^2_{(A_t+B_K^t)^{-1}} \leq ||x||^2_{(A_t+B_{K-1}^t)^{-1}} \leq ||x||^2_{A_t^{-1}} = ||x||^2_{(\lambda I_N + N_a(t)x_a x_a^T)^{-1}}$.

□

Lemma 12. $\forall t > 0, \forall a \in [K], \forall y \in \mathbb{R}^N$, $||y||_{(\hat{V}_t^\lambda)^{-1}} \leq ||y||/\sqrt{N_a(t)||x_a||^2 + \lambda}$. (Lemma 2 in the paper)

Proof. Using successively Lemma 11, Sherman-Morrison formula and Cauchy-Schwarz inequality:

$$
\begin{aligned}
||y||^2_{(\hat{V}_t^\lambda)^{-1}} \leq ||y||^2_{(\lambda I_N + N_a(t)x_a x_a^\top)^{-1}} &= \frac{||y||^2}{\lambda} - \frac{\lambda^{-2}N_a(t)(< y, x_a >)^2}{1 + \lambda^{-1}N_a(t)||x_a||^2}\\
&\leq \frac{||y||^2}{\lambda} - \frac{\lambda^{-2}N_a(t)||y||^2||x_a||^2}{1 + \lambda^{-1}N_a(t)||x_a||^2} = \frac{||y||^2}{\lambda + N_a(t)||x_a||^2}
\end{aligned}
$$

□

Lemma 13. *Let $T^* : [K] \times (0,1) \times \mathbb{N}^* \to \mathbb{R}^{*+}$ be a function that is nondecreasing in t, and $\mathcal{I}_t$ the set of pulled arms at time t. Let $\mathcal{E}$ be an event such that for all $t < \tau_\delta, \delta \in (0,1)$, $\exists a_t \in \mathcal{I}_t, N_{a_t}(t) \leq T^*(a_t, \delta, t)$. Then it holds on the event $\mathcal{E}$ that $\tau_\delta \leq T(\mu, \delta)$ where*

$$T(\mu,\delta) \triangleq \inf\left\{u \in \mathbb{R}^{*+} : u > 1 + \sum_{a=1}^{K} T^*(a,\delta,u)\right\}.$$

(Lemma 6 in the paper)

Proof. Let us denote $T \in \mathbb{N}^*$. Let us study $\min(\tau_\delta, T)$, because $\min(\tau_\delta, T) < T \implies \tau_\delta < T$. On event $\mathcal{E}$:

$$\begin{aligned}
\min(\tau_\delta, T) &= 1 + \sum_{t \leq T} \mathbb{1}(t < \tau_\delta) \leq 1 + \sum_{t \leq T} \mathbb{1}(\exists a_t \in \mathcal{I}_t, N_{a_t}(t) \leq T^*(a_t, \delta, t)) \text{ (using definition of } T^*\text{, and } \mathcal{E} \text{ holds)} \\
&= 1 + \sum_{t \leq T} \sum_{m=1}^{t} \mathbb{1}(\exists a_t \in \mathcal{I}_t, N_{a_t}(t) = m \wedge m \leq T^*(a_t, \delta, t)) \text{ (using } \forall a \in [K], N_a(t) \in [t] \wedge \forall a \in \mathcal{I}_t, N_a(t) > 0) \\
&\leq 1 + \sum_{m=1}^{T} \sum_{t=m}^{T} \sum_{a \in [K]} \mathbb{1}(a \in \mathcal{I}_t) \mathbb{1}(N_a(t) = m \wedge m \leq T^*(a, \delta, t)) \text{ (using the union bound on pulled arms)} \\
&= 1 + \sum_{a \in [K]} \sum_{m=1}^{T} \sum_{t=m}^{T} \mathbb{1}(a \in \mathcal{I}_t) \mathbb{1}(N_a(t) = m) \mathbb{1}(m \leq T^*(a, \delta, t)) \\
&\leq 1 + \sum_{a \in [K]} \sum_{m=1}^{T} \left[\sum_{t=m}^{T} \mathbb{1}(a \in \mathcal{I}_t) \mathbb{1}(N_a(t) = m)\right] \mathbb{1}(m \leq T^*(a, \delta, T)) \text{ (since } T^* \text{ is nondecreasing in } t) \\
&\leq 1 + \sum_{a \in [K]} \sum_{m=1}^{T} 1 \times \mathbb{1}(m \leq T^*(a, \delta, T)) \leq 1 + \sum_{a \in [K]} T^*(a, \delta, T)
\end{aligned}$$

Choosing any T that satisfies $1 + \sum_{a \in [K]} T^*(a, \delta, T) < T$ yields $\min(\tau_\delta, T) < T$ and therefore $\tau_\delta \leq T$. The smallest possible such T is

$$T(\mu,\delta) \triangleq \inf\left\{u \in \mathbb{R}^{*+} : u > 1 + \sum_{a \in [K]} T^*(a,\delta,u)\right\}.$$

□