Dynamic Regret Analysis for Online Tracking of Time-varying Structural Equation Model Topologies

Bakht Zaman, Student Member, IEEE, Luis Miguel Lopez Ramos, Member, IEEE, and Baltasar Beferull-Lozano, Senior Member, IEEE

WISENET Center, Department of ICT, University of Agder, Grimstad, Norway

Abstract—Identifying dependencies among variables in a complex system is an important problem in network science. Structural equation models (SEM) have been used widely in many fields for topology inference, because they are tractable and incorporate exogenous influences in the model. Topology identification based on static SEM is useful in stationary environments; however, in many applications a time-varying underlying topology is sought. This paper presents an online algorithm to track sparse time-varying topologies in dynamic environments and most importantly, performs a detailed analysis on the performance guarantees. The tracking capability is characterized in terms of a bound on the dynamic regret of the proposed algorithm. Numerical tests show that the proposed algorithm can track changes under different models of time-varying topologies.

I. INTRODUCTION

Time series are generated and observed in many applications. Using multiple time series data from a complex system, identifying a structure explaining dependencies (connections) among variables is a well-motivated problem in many fields [1]. Such a networked structure may offer insights about the system dynamics and can assist in inference tasks such as prediction, event detection, and signal reconstruction [2], [3], [4].

There are different models and approaches that are extensively used in topology identification in certain applications: see, e.g., [5], [6], [3], and references therein. Among these models, structural equation model (SEM) is a popular model [7]: this is mainly due to its tractability and the ability to identify directed relations by means of the inclusion of exogenous variables, which are naturally available in many applications. These exogenous variables represent influences that do not depend on the (endogenous) variables in the model, and their inclusion contributes to the model identifiability [8]. Static SEMs have been applied to topology identification problems in various fields, e.g., gene regulatory network discovery from gene expression data [9]. However, static SEM cannot capture topology changes if the underlying dynamics are nonstationary and each observation is obtained at instants relatively spaced in time, which occurs in various applications.

In time-varying environments, a dynamic SEM can be applied [10]. A dynamic SEM is considered in [11] to track information cascades of popular news topics over social networks, which are assumed to have sparse dynamic topologies. In the same work, several online algorithms are presented, but not supported by any performance guarantees, so that their tracking capabilities are not theoretically characterized. In [12], an online algorithm for tracking dynamic topologies is proposed where the exogenous input is not fully known, also without convergence guarantees.

In this paper, an online algorithm to track the changes in dynamic SEM topologies in the lines of [11] is described and its dynamic regret is analyzed, to theoretically characterize its tracking capabilities. The dynamic regret measures the cumulative difference between the cost function evaluated at the estimates and the cost function evaluated at a sequence of time-varying optimal solutions. Specifically, we provide a bound on the dynamic regret that depends on easily measurable properties of the data, the algorithm hyperparameters, and a metric of how much the model varies along time.

The rest of the paper is organized as follows: Sec. II contains the model, problem formulation, and the derivation of the algorithm. Sec. III establishes the dynamic regret bound, including its formal proofs. Numerical results are presented in Sec. IV and Sec. V concludes the paper.

II. MODEL AND PROBLEM FORMULATION

Consider a networked system with $N$ nodes, indexed by $i$. At each time frame indexed by $t$, a number $C$ of interactions (frequently denoted as contagions) indexed by $c$ are observed in the system, with $y_{ic}^{t}$ denoting the intensity of the $c$-th contagion in node $i$ at time $t$. Also, let $x_{ic}$ denote the susceptibility of node $i$ to external influence (infection) by contagion $c$. The dynamic linear structural equation model (SEM) is given by [11]:

$$y_{ic}^{t} = \sum_{j=1,j\neq i}^{N} a_{ij}^{t}b_{jc}^{t} + b_{ic}^{t}x_{ic} + e_{ic}^{t}, \quad (1)$$

for $i = 1, \ldots, N$, $c = 1, \ldots, C$, $t = 1, \ldots, T$, where the coefficients $a_{ij}^{t}$ are the time-varying SEM parameters that encode the topology of the network, $b_{ic}^{t}$ quantifies the level of influence of external sources on node $i$, and $e_{ic}^{t}$ denotes the measurement errors and un-modeled dynamics. By defining $y_{c}^{t} = [y_{1c}^{t}, \ldots, y_{Nc}^{t}]^{\top} \in \mathbb{R}^{N}$, $x_{c} = [x_{1c}, \ldots, x_{Nc}]^{\top} \in \mathbb{R}^{N}$, $B^{t} = \text{diag}(b^{t}) \in \mathbb{R}^{N \times N}$ with $b^{t} = [b_{11}^{t}, \ldots, b_{NN}^{t}]^{\top}$, and
The matrix $A^t \in \mathbb{R}^{N \times N}$ can be seen as a time-varying adjacency matrix for an SEM-based network. The observations for all contagions can be collected in a matrix by defining $Y^t = [y_1^t, \ldots, y_C^t] \in \mathbb{R}^{N \times C}$, $X = [x_1, \ldots, x_C] \in \mathbb{R}^{N \times C}$, and $E^t = [e_1^t, \ldots, e_C^t] \in \mathbb{R}^{N \times C}$. The dynamic SEM takes the following form:

$$Y^t = A^t Y^t + B^t X + E^t.$$  (3)

The parameter $\gamma \in (0, 1]$ is a forgetting factor that regulates how much past information influences the solution at time $t$, and $\lambda$ is the sparsity-promoting regularization parameter. The constraint $a_{ii} = 0$ eliminates any component of the trivial solution $A = I$. The constraint $b_{ij} = 0$ guarantees a diagonal $B$, meaning that external sources for a certain node do not affect any other node $j \neq i$. Dealing with constraints can be easily avoided if we rewrite the objective including only the nonzero elements of the matrices. We can rewrite $f_t(A, B)$ as:

$$f_t(A, B) = \frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T} \gamma^{t-r} \left\| y_{i}^T - a_{i}^T Y_{i}^T - b_{i} x_{i}^T \right\|_F^2,$$  (6a)

$$= \frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T} \gamma^{t-r} \left\| y_{i}^T - [a_{i}^T b_{i}] \left( Y_{i}^T \right)_{x_{i}} \right\|_F^2.$$  (6b)

where $y_{i}^T$ is the $i$-th row of $Y^T$, $x_{i}$ is the $i$-th row of $X$, $a_{i}^T$ is the $i$-th row of $A$ without $i$-th entry, and $Y_{i}^T$ is obtained by removing the $i$-th row from $Y^T$.

Further, we can define $v_i = [a_{i}^T b_{i}]^T$ and $Z_{i}^T \triangleq [(Y_{i}^T)^T x_{i}]^T$ to rewrite (6a):

$$f_t(A, B) = \frac{1}{2} \sum_{i=1}^{N} \sum_{t=1}^{T} \gamma^{t-r} \left\| y_{i}^T - v_{i}^T Z_{i}^T \right\|_F^2,$$  (7)

Note that $f_t$ in (6a) is separable across $i$ (nodes), so that

$$f_t(A, B) = \sum_{i=1}^{N} f_i(v_i),$$  (8a)

where

$$f_i(v_i) = \frac{1}{2} \sum_{t=1}^{T} \gamma^{t-r} \left\| y_{i}^T - (Z_{i}^T)^T v_{i} \right\|_2^2.$$  (8b)

Similarly, upon defining

$$\Omega_i(v_i) \triangleq \lambda \left\| a_{-i} \right\|_1,$$  (9)

the regularization function is also separable across the rows of $A$, as $\|\text{vec}(A)\|_1 = \sum_{i=1}^{N} \Omega_i(v_i)$.

In the next subsection, the online proximal gradient algorithm in [13] will be applied to solve (5) leveraging the separability we just presented. Before presenting the algorithm, we re-write $f_i(v_i)$ in a form that will simplify the computation of its gradient. By expanding (8b) and ignoring terms not dependent on $v_i$:

$$f_i(v_i) \propto \frac{1}{2} \sum_{t=1}^{T} \gamma^{t-r} \left[ v_{i}^T Z_{i}^T(Z_{i}^T)^T v_{i} - 2y_{i}^T(Z_{i}^T)^T v_{i} \right],$$

the gradient of $f_i(v_i)$ is given by

$$\nabla v_i f_i(v_i) = \Phi_{Z_i} v_{i} - r_{i},$$  (10a)

where

$$\Phi_{Z_i} \triangleq \sum_{r=1}^{t} \gamma^{t-r} Z_{i}^T(Z_{i}^T)^T$$  (10b)

and

$$r_{i} \triangleq \sum_{t=1}^{T} \gamma^{t-r} Z_{i}^T(y_{i}^T)^T.$$  (10c)

Note that the variables defined in the latter two expressions can be computed recursively, as will be expressed in the tabulated algorithm (lines 5 and 6).

**A. Proximal online gradient algorithm**

The update of the proximal online gradient descent algorithm [13], applied to the $i$-th portion of the separable problem presented in the previous section, yields

$$v_i[t+1] = \text{prox}_{\alpha_i}^{\gamma_i} (g_i^*(v_i[t])), \quad \alpha > 0, \quad g_i^*(u) = \Phi_{Z_i} u - \alpha \nabla u f_i(u), \quad \text{prox}_{\alpha_i}^{\gamma_i}(w) = \arg \min_{s \in \text{dom} \Psi} \left[ \Psi(s) + \frac{1}{2 \alpha} \left\| s - w \right\|_2^2 \right].$$  (12)

From the definition of $v_i$ and (9), it becomes clear that

$$\text{prox}_{\alpha_i}^{\gamma_i}(s) = \left[ S_{\alpha \lambda}(s) : 1_{N-1}^T s \right]_N^T$$  (13)

with $S_{\alpha \lambda}(w)$ denoting the standard soft-thresholding operator. The complete procedure is presented in **Algorithm 1**. Observe that the step size $\alpha$ is required to be small enough, specifically $\alpha < 1/L_f$ where $\lambda_{\text{max}}(\Phi_{Z_i}) \leq L_f, \forall i, t.$
Algorithm 1 Online algorithm for tracking dynamic SEM-based Topologies

Input: \( \gamma, \lambda, \alpha \in (0, 1/L_f), \{Y^i\}_{i=1}^T, X \)

Output: \( \{A[t]\}_{t=1}^T, \{B[t]\}_{t=1}^T \)

Initialization: \( v_i[1] = 0_{N \times 1}, \Phi_{Z_i} = 0_{N \times N}, r_i^0 = 0_{N \times 1}, i = 1, \ldots, N \)

for \( t = 1, 2, \ldots, T \) do

1. Receive data \( Y^t \)
2. for \( i = 1, 2, \ldots, N \) do
3. \[ Z_i^t = \left( (Y_i^t)_n \right)^\top \left( (x_i^t)_n \right)^\top \]
4. \[ \Phi_{Z}^t = \gamma \Phi_{Z}^{t-1} + Z_i(z_i^t)^\top \]
5. \[ r_i^t = \gamma r_i^{t-1} + Z_i(y_i^t)^\top \]
6. \[ \nabla_v f_i^t(v_i[t]) = \Phi_{Z}^t v_i[t] - r_i^t \]
7. \[ v_i[t+1] = \max_{y_i} \alpha \nabla_v f_i^t(v_i[t]) \]
end for

Form \( A[t] \) and \( B[t] \) from \( v_i[t], i = 1, \ldots, N \)
end for

III. DYNAMIC REGRET ANALYSIS

The performance of online algorithms is evaluated by means of the regret, which is the difference in performance between the online algorithm and a solution which can be computed based on the data in hindsight. The regret measure can be static or dynamic. In the case of the static regret, the best comparator implicitly assumes a stationary model. Therefore, the static regret cannot express the tracking performance of an online algorithm in dynamic environments, where the generating parameters are time-varying. To characterize the tracking performance of online algorithms, the dynamic regret [14] is used, which results from comparing the online algorithm against an optimal sequence of time-varying hindsight solutions. Specifically, upon defining \( h_i(A[t], B[t]) = f_i(A[t], B[t]) + \Omega(A[t]) \), the dynamic regret is given by:

\[
R_d[T] = \sum_{t=1}^T \left[ h_i(A[t], B[t]) - h_i(A^*[t], B^*[t]) \right].
\]  

(14)

with \( (A^*[t], B^*[t]) \) representing the estimate produced by aclairvoyant that knows \( h_i(\cdot) \) in advance (in contrast, the online algorithm does not have access to \( h_i(\cdot) \) while producing \( (A[t], B[t]) \)). Using (6a), the above expression can be written as:

\[
R_d[T] = \sum_{t=1}^T \sum_{i=1}^N \left[ f_i^t(v_i[t]) + \Omega^i(v_i[t]) - f_i^t(v_i^*[t]) - \Omega^i(v_i^*[t]) \right]
\]

\[
= \sum_{i=1}^N \sum_{t=1}^T \left[ h_i^t(v_i[t]) - h_i^t(v_i^*[t]) \right] = \sum_{i=1}^N R_d^i[T],
\]

where \( h_i^t(v_i[t]) \triangleq f_i^t(v_i[t]) + \Omega^i(v_i[t]), v_i^*[t] \triangleq \arg \min_{v_i} f_i^t(v_i) + \Omega^i(v_i), \) and \( R_d^i[T] \triangleq \sum_{t=1}^T [h_i^t(v_i[t]) - h_i^t(v_i^*[t])] \). Observe that the regret expression is separable across index \( i \) (nodes). Thus, for the sake of simplicity, we derive the regret for the \( i \)-th node, i.e., \( R_d^i[T] \). The total regret will be obtained by adding the individual regret expressions.

We define the path length for each subproblem (corresponding to each node \( i \)) as:

\[
W_i[T] \triangleq \sum_{t=2}^T \| v_i^t[t] - v_i^t[t-1] \|_2,
\]

(15)

which represents the aggregated variations in the consecutive optimal solutions. In this work, the following assumptions are considered:

A1. Bounded process: There exists \( B_{xy} \) such that \( \| y_{ic}^t \|_2 \leq B_{xy} \) for all \( i, c, t \).

A2. Strong convexity: Each function \( f_i^t \) is \( \beta \)-strongly convex, i.e., \( \lambda_{\min}(\Phi_{Z_i}) \geq \beta > 0 \).

A3. Lipschitz smoothness: Each function \( f_i^t \) is \( L_f \)-Lipschitz smooth, i.e., \( \lambda_{\max}(\Phi_{Z_i}) \leq L_f, \forall i, t \).

A4. Bounded variations of the optimal solution: The distance between two consecutive optimal solution is bounded, i.e.,

\[
\| v_i^t[t] - v_i^t[t+1] \|_2 \leq \beta, d \geq 0, \forall t, i.
\]

(16)

These above assumptions are standard in the literature. Assumption A1 does not entail any loss of generality and is satisfied in most real-world applications. Next, we present an upper bound on the dynamic regret.

Theorem 1: The individual dynamic regret of Algorithm 1 for a node \( i \) is given by:

\[
R_d^i[T] = D_h (\| v_i^*[1] \|_2 + W_i[T]),
\]

where

\[
D_h \triangleq \frac{1}{\alpha \beta} \left( \frac{B_{xy} C \sqrt{N}}{1 - \gamma} \left( 1 + \frac{L_f}{\beta} \right) + \lambda \sqrt{N - 1} \right),
\]

(18)

under assumptions A1, A2, A3, and A4.

Proof: Since \( h_i^t \) is convex, we have by definition that:

\[
h_i^t(v_i^*[t]) \geq h_i^t(v_i[t]) + (\nabla h_i^t(v_i[t]))^\top (v_i^*[t] - v_i[t]),
\]

(19)

\[ \forall v_i[t], v_i^*[t], \nabla h_i^t(v_i[t]) \text{ denotes a subgradient of } h_i^t(v_i[t]) \]
given by \( \nabla h_i^t(u) = \nabla f_i(u) + \nabla \Omega^i(u) \) with \( \nabla \Omega^i(u) \in \partial \Omega^i(u) \). Rearranging and summing the above inequality from \( t = 1 \) to \( T \), we have

\[
\sum_{t=1}^T [h_i^t(v_i[t]) - h_i^t(v_i^*[t])] \leq \sum_{t=1}^T (\nabla h_i^t(v_i[t]))^\top (v_i[t] - v_i^*[t])
\]

\[
\leq \sum_{t=1}^T \| \nabla h_i^t(v_i[t]) \|_2 \cdot \| v_i[t] - v_i^*[t] \|_2,
\]

(20)

where the second inequality follows from the Cauchy-Schwarz inequality. Next, we derive a bound on \( \| \nabla h_i^t(v_i[t]) \|_2 \). Note first that it holds that

\[
\| \nabla h_i^t(v_i[t]) \|_2 \leq \| \nabla f_i^t(v_i[t]) \|_2 + \| \nabla \Omega^i(v_i[t]) \|_2.
\]

(21)

Thus, we have to prove that \( \| \nabla f_i^t(v_i[t]) \|_2 \) and \( \| \nabla \Omega^i(v_i[t]) \|_2 \) are bounded \( \forall v_i[t] \). First, we prove that \( \| \nabla f_i^t(v_i[t]) \|_2 \) is...
bounded. To this end, from (10a), using the triangular inequality, the spectral radius of $\Phi_{Z_t}^i$, and assumption A3, we obtain the following:

$$
\|\nabla f_t^i(v_t[t+1])\|_2 = \|\Phi_{Z_t}^i v_t[t+1] - r_t^i\|_2 \\
\leq \|\Phi_{Z_t}^i v_t[t+1]\|_2 + \|r_t^i\|_2 \\
\leq \max_{i} \|\Phi_{Z_t}^i\| v_t[t+1] + \|r_t^i\|_2 \\
\leq L_f \|v_t[t+1]\|_2 + \|r_t^i\|_2.
$$

(22)

We need to derive a bound on $\|v_t[t+1]\|_2$ and $\|r_t^i\|_2$. First, we derive a bound on $\|r_t^i\|_2$. From the definition of $r_t^i$ in (10c), and using assumption A1, we obtain the bound as follows:

$$
\|r_t^i\|_2 = \left\| \sum_{t=1}^T \gamma^{t-t} Z_t^i (y_t^i)^T \right\|_2 \leq \sum_{t=1}^T \gamma^{t-t} B_{xy} \|1_N \times C\|_2 \\
= B_{xy} \left\| \sum_{t=1}^T \gamma^{t-t} C1_N \right\|_2 = B_{xy} \sum_{t=1}^T \gamma^{t-t} \|1_N\|_2 \\
\leq \frac{B_{xy} C \sqrt{N}}{1 - \gamma} = \frac{B_{xy} C \sqrt{N}}{\mu},
$$

(23)

where $\mu \triangleq 1 - \gamma$. Thus, we have derived a bound on $\|r_t^i\|_2$. To derive an upperbound on $\|v_t[t+1]\|_2$, from the update expression of the algorithm, and using assumption A2, we have that:

$$
\|v_t[t+1]\|_2 \leq \|v_t[t] - \alpha \nabla f_t^i(v_t[t])\|_2 \\
= \|v_t[t] - \alpha (\Phi_{Z_t}^i v_t[t] - r_t^i)\|_2 \\
= \|(I - \alpha \Phi_{Z_t}^i) v_t[t] + \alpha r_t^i\|_2 \\
\leq \lambda_{\max} \|I - \alpha \Phi_{Z_t}^i\| v_t[t] + \alpha \|r_t^i\|_2 \\
= (1 - \alpha \lambda_{\min}(\Phi_{Z_t}^i)) v_t[t] + \alpha \|r_t^i\|_2 \\
\leq (1 - \alpha \beta) \|v_t[t]\|_2 + \alpha \|r_t^i\|_2.
$$

(24)

Substituting the bound on $r_t^i$ from (23) in the above inequality, we obtain:

$$
\|v_t[t+1]\|_2 \leq (1 - \alpha \beta) \|v_t[t]\|_2 + \alpha \frac{B_{xy} C \sqrt{N}}{\mu} \\
\leq \delta \|v_t[t]\|_2 + \alpha \frac{B_{xy} C \sqrt{N}}{\mu}.
$$

By recursive substitution in the above inequality:

$$
\|v_t[t+1]\|_2 \leq \delta (\|v_t[t-1]\|_2 + \alpha \frac{B_{xy} C \sqrt{N}}{\mu}) + \alpha \frac{B_{xy} C \sqrt{N}}{\mu} \\
\leq \delta^2 \|v_t[t-2]\|_2 + \alpha \frac{B_{xy} C \sqrt{N}}{\mu} (\delta^2 + \delta + 1) \leq \ldots \\
\leq \delta^k \|v_t[k+1]\|_2 + \alpha \frac{B_{xy} C \sqrt{N}}{\mu} \sum_{i=0}^{k-1} \delta^i,
$$

where $1 \leq k \leq t$. For $k = t$, the above inequality becomes

$$
\|v_t[t+1]\|_2 \leq \delta^t \|v_t[1]\|_2 + \alpha \frac{B_{xy} C \sqrt{N}}{\mu} \sum_{i=0}^{t-1} \delta^i \\
= \frac{\alpha B_{xy} C \sqrt{N}}{\mu} \left\{ 1 - \delta^t \right\} \leq \frac{\alpha B_{xy} C \sqrt{N}}{\mu} \frac{1}{\alpha \beta} \\
= \frac{B_{xy} C \sqrt{N}}{\mu \beta}.
$$

(26)

By substituting the bounds from (26) and (23) into (22), we obtain the bound on $\|\nabla f_t^i(v_t[t+1])\|_2$ as follows:

$$
\|\nabla f_t^i(v_t[t+1])\|_2 \leq \frac{L_f B_{xy} C \sqrt{N}}{\mu \beta} + \frac{B_{xy} C \sqrt{N}}{\mu} \\
= \frac{B_{xy} C \sqrt{N}}{\mu} \left(1 + \frac{L_f}{\beta}\right).
$$

(27a)

(27b)

To prove that $\|\nabla \Omega^i(v_t[t])\|_2$ is bounded $\forall v_t[t]$, first we compute the Lipschitz continuity parameter of $\Omega^i$, i.e., $L_\Omega$ and then apply the result in [15, Lemma 2.6], which establishes that all the subgradients of a function are bounded by its Lipschitz continuity parameter. To find $L_\Omega$, let $a' \triangleq [a'm_n]^T, b' \triangleq [b' n_n]^T, a, b \in \mathbb{R}^{N-1}, m, n \in \mathbb{R}$. By the triangular inequality and the reverse triangular inequality, we have that:

$$
\|\Omega^i(a') - \Omega^i(b')\|_2 = \|\lambda|a| - \lambda |b|\|_1 \\
= \lambda \sum_{i=1}^{N-1} \|a_i - b_i\|_1 \\
\leq \lambda \sum_{i=1}^{N-1} \lambda \|a_i - b_i\|_1 \\
\leq \lambda \|a - b\|_1 \leq \lambda \sqrt{N-1} \|a - b\|_2 \\
\leq\lambda \sqrt{N-1} \|a' - b'\|_2.
$$

Thus, we have that $L_\Omega = \lambda \sqrt{N-1}$. Substituting these bounds in (21), we have

$$
\frac{\|\nabla (f_t^i(v_t))\|_2 + \|\nabla \Omega^i(v_t)\|_2}{\|v_t\|_2} \leq \frac{B_{xy} C \sqrt{N}}{\mu \beta} \left(1 + \frac{L_f}{\beta}\right) + \lambda \sqrt{N-1}.
$$

(28)

Substituting the above bound in (20), we obtain

$$
\sum_{t=1}^{T} [h_t^i(v_t[t]) - h_t^i(v_t^*)] \\
\leq \sum_{t=1}^{T} \left( \frac{B_{xy} C \sqrt{N}}{\mu} \left(1 + \frac{L_f}{\beta}\right) + \lambda \sqrt{N-1} \right) \|v_t[t] - v_t^*[t]\|_2 \\
= \left( \frac{B_{xy} C \sqrt{N}}{\mu} \left(1 + \frac{L_f}{\beta}\right) + \lambda \sqrt{N-1} \right) \sum_{t=1}^{T} \|v_t[t] - v_t^*[t]\|_2.
$$

(29)

Next, we derive a bound on $\sum_{t=1}^{T} \|v_t[t] - v_t^*[t]\|_2$. To this end, the first step is to prove the following result:

$$
\|v_t[t+1] - v_t^*[t]\|_2 \leq \rho \|v_t[t] - v_t^*[t]\|_2,
$$

(30)
where $\rho = 1 - \alpha \beta$. To this end, squaring the l.h.s. of (30) and by definition of $v_i[t + 1]$, we have

$$
\|v_i[t + 1] - v_i^*[t]\|^2_2 = \left\|\alpha a_i[t + 1] - \alpha a_i^*[t] + b_i[t + 1] - b_i^*[t]\right\|^2_2 \\
\leq \left\|\alpha a_i[t + 1] - \alpha a_i^*[t]\right\|^2_2 + \|b_i[t + 1] - b_i^*[t]\|^2_2 \\
= \left\|\alpha \text{prox}_{\|\cdot\|_2}(a_i[t] - \alpha \nabla a_i f_i(a_i[t])) - \text{prox}_{\|\cdot\|_2}(a_i^*[t] - \alpha \nabla a_i f_i(a_i^*[t]))\right\|^2_2 \\
+ (b_i[t] - \alpha \nabla b_i f_i(b_i[t]) - (b_i^*[t] - \alpha \nabla b_i f_i(b_i^*[t])))^2 \\
\leq \left(\|a_i[t] - a \nabla a_i f_i(a_i[t]) - a \nabla a_i f_i(a_i^*[t])\|^2_2 + \|b_i[t] - \alpha \nabla b_i f_i(b_i[t]) - (b_i^*[t] - \alpha \nabla b_i f_i(b_i^*[t]))\|^2_2\right) \\
= \|v_i[t] - v_i^*[t]\|^2_2 + \frac{\alpha^2}{L_f} \|\nabla v_i f_i(v_i[t]) - \nabla v_i f_i(v_i^*[t])\|^2_2 \\
- 2\alpha \left(\frac{\beta L_f}{L_f + \beta} \|v_i[t] - v_i^*[t]\|^2_2 + \frac{1}{L_f + \beta} \|\nabla v_i f_i(v_i[t]) - \nabla v_i f_i(v_i^*[t])\|^2_2\right),
$$

where the above inequality is implied by assumptions A2 and A3. Given that $\alpha \in (0, 1/L_f)$, using assumption A2, and by further simplifications, we have:

$$
\|v_i[t + 1] - v_i^*[t]\|^2_2 \\
\leq \left(1 - \frac{2\alpha \beta L_f}{L_f + \beta}\right) \|v_i[t] - v_i^*[t]\|^2_2 + \left(\alpha^2 - \frac{2\alpha}{L_f + \beta}\right). \\
\|\nabla v_i f_i(v_i[t]) - \nabla v_i f_i(v_i^*[t])\|^2_2 \\
= \left(1 - \frac{2\alpha \beta L_f}{L_f + \beta}\right) \|v_i[t] - v_i^*[t]\|^2_2 - \left(\frac{2\alpha}{L_f + \beta} - \alpha^2\right). \\
\|\nabla v_i f_i(v_i[t]) - \nabla v_i f_i(v_i^*[t])\|^2_2 \\
\leq \left(1 - \frac{2\alpha \beta L_f}{L_f + \beta}\right) \|v_i[t] - v_i^*[t]\|^2_2 - \beta^2 \left(\frac{2\alpha}{L_f + \beta} - \alpha^2\right). \\
\|v_i[t] - v_i^*[t]\|^2_2 \\
= \left(1 - \frac{2\alpha \beta L_f}{L_f + \beta} - \frac{2\alpha^2}{L_f + \beta} + \alpha^2 \beta^2\right) \|v_i[t] - v_i^*[t]\|^2_2 \\
= (1 - 2\alpha \beta + \alpha^2 \beta^2) \|v_i[t] - v_i^*[t]\|^2_2 \\
= \rho^2 \|v_i[t] - v_i^*[t]\|^2_2,
$$

where $\rho \triangleq 1 - \alpha \beta$. Taking square root on both sides of the above inequality yields (30).

Next, we show that

$$
\sum_{i=1}^{T} \|v_i[t] - v_i^*[t]\|^2_2 \leq \frac{1}{1 - \rho} \|v_i[1] - v_i^*[1]\|^2_2 + W_i[T].
$$

To prove the above expression, consider the cumulative gap

$$
\sum_{t=2}^{T} \|v_i[t] - v_i^*[t]\|^2_2 \leq \sum_{t=2}^{T} \|v_i[t] - v_i^*[t-1] + v_i^*[t-1] - v_i^*[t]\|^2_2 \\
\leq \sum_{t=2}^{T} \|v_i[t] - v_i^*[t-1]\|^2_2 + \|v_i^*[t] - v_i^*[t-1]\|^2_2 \\
= \sum_{t=1}^{T} \|v_i[t + 1] - v_i^*[t]\|^2_2 + W_i[T] \\
\leq \sum_{t=1}^{T} \rho \|v_i[t] - v_i^*[t]\|^2_2 + W_i[T].
$$

Adding $\|v_i[1] - v_i^*[1]\|^2_2$ on both sides of the above inequality results in:

$$
\sum_{i=1}^{T} \|v_i[t] - v_i^*[t]\|^2_2 \leq \sum_{i=1}^{T} \rho \|v_i[t] - v_i^*[t]\|^2_2 + \|v_i[1] - v_i^*[1]\|^2_2 + W_i[T].
$$

By rearranging terms in the above inequality, we obtain the result in (33). Thus, we can substitute $\sum_{t=1}^{T} \|v_i[t] - v_i^*[t]\|^2_2$ with its bound from (33) into (29) and note that $v_i[1] = 0 \times N \times 1$ in Algorithm 1. This completes the proof.

**Remarks.** The bound on the total dynamic regret is given by:

$$
R_d[T] = D_h \sum_{i=1}^{N} \|v_i[1]\|^2_2 + W_i[T],
$$

where $D_h$ is defined in (18). Notice that this means that the bound on the dynamic regret is a function of the parameters of the data and the parameters of the algorithm. Moreover, for a sublinear path length, the dynamic regret of the proposed algorithm is sublinear.

**IV. Numerical Results**

In this section, the performance of the algorithm is analyzed by presenting numerical tests. The experimental results are based on synthetic data.

To generate the matrices $A^t$, a binary adjacency matrix $A_{\text{binary}}$ is generated according to an Erdős-Rényi model with edge probability $p_e$. No self-loops are considered, i.e., the diagonal entries of $A_{\text{binary}}$ are zero. Two models are considered in the simulations: a) smooth-transition model and b) non-smooth transition model. In the smooth-transition model, the nonzero elements of $A^t$ follow the pattern of the 1’s in $A_{\text{binary}}$. For $t = 1$, the nonzero elements of $A^t$ take one of the following four functions via random selection: i) $a1(t) = 0.5 + 0.5 \sin(0.1t)$, ii) $a2(t) = 0.5 + 0.5 \cos(0.1t)$, iii) $a3(t) = \exp(-0.01t)$, and iv) $a4(t) = 0$. For $t > 1$, the elements of $A^t$ evolve according to the function selected initially by evaluating the functions for $t$. In the non-smooth transition model, a static model for $A^t$ is considered for $t < T/2$. The nonzero elements of $A^t$ are drawn one time
which is a metric of how much the model parameters vary in a time interval. When the path length is sublinear in time, the dynamic regret of the algorithm becomes sublinear, meaning that the online algorithm enjoys a performance comparable to the optimal offline estimator. The tracking capabilities of the algorithm have been numerically validated for a time-varying scenario under two different assumptions on the model variations, namely a smooth-transition and an abrupt-transition model.

V. CONCLUSION

An online algorithm for tracking dynamic SEM-based topologies is presented in this paper. A bound was derived on the dynamic regret (a much better metric than static regret for time-varying scenarios) of the proposed algorithm. This bound is a function of the numeric properties of the data that are easy to obtain, parameters of the algorithm, and the path length,