Robustness of Graph Topology Learning with Smooth Signals under Partial Observations

Hoang-Son Nguyen
Oregon State University
nguyhoa3@oregonstate.edu

Hoi-To Wai
The Chinese University of Hong Kong
htwai@se.cuhk.edu.hk

***Abstract*—Recently, many sophisticated algorithms have been proposed for graph topology learning from partial observations. Most of them have relied on advanced structures such as low-rankness and sparsity, but would otherwise require the number of unobserved nodes to be significantly smaller than the graph size. The aim of this ongoing work is to demonstrate theoretically that simple graph topology learning methods are implicitly robust to partial observations of low pass filtered graph signals. We achieve this result through extending the RIP property for the Dirichlet energy function. We show that smoothness-based graph learning formulation on partial observations is able to learn the ground truth graph topology corresponding to the observed nodes.**

I. Introduction

This ongoing work studies the problem of graph topology learning from a set of graph signal observations. As a crucial first step for graph data analysis, etc., graph topology learning has received attention from signal processing, control, and machine learning communities [1]–[4]. A typical setup in the literature is to consider *full observation* data where the graph signals on each node are recorded (simultaneously). However, for graphs with large number of nodes or even open systems, obtaining full observation of graph signals can be challenging. For social networks, this requires each individual to report their states within a limited time interval; for biological or physical networks, this requires estimating the states of each components. As such, we are often restricted to accessing *partially observed* graph signals, where the states of a subset of nodes become unobservable. These unobserved nodes are *hidden* from us whose existence may not even be known.

Although the states of hidden nodes are not observed, these nodes may still influence the observed nodes in the network system. As pioneered by [5], recent works have considered sophisticated methods for graph topology learning while accounting for the influence of these hidden nodes. Importantly, [5] showed that the unknown influence signals are low rank if the number of hidden nodes is lower than the number of observed nodes. This inspired recent works to propose graph topology learning methods that are aware of the hidden nodes for Gaussian graphical model inference [5], stationary graph signals via spectral template [6], smooth graph signals [6], etc. The major drawback of these works is that they generally require the number of hidden nodes to be *significantly lower* than the number of observed nodes. The new graph learning formulations also require additional complexity with hyperparameter tuning.

Our aim is to explore an alternative strategy to graph topology learning with partial observation which is *agnostic* to the existence of hidden nodes. We study a 'naive' approach of directly applying graph topology learning methods for full observations such as [1], [2] on partially observed graph signals. In particular, we concentrate on analyzing the *robustness* of the smoothness based graph learning objective, i.e., the Dirichlet energy of graph signals, under partial observations. We will present the following findings:

- If the graph signals are generated from a low pass filtering process, then such a naive approach is guaranteed to learn a partial graph topology similar to one taken from learning with full observation.
- We show the above theoretical result through utilizing a new perspective on the restricted isometry property (RIP) for quadratic forms of the sampled graph Laplacian.

In other words, for low pass graph signals, it suffices to directly apply [1] to infer the partial graph topologies, instead of relying on further sophistication in [5], [6].

II. Main Result

Setup & Notations. Consider an N-node weighted undirected connected simple graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \ldots, N\}$ and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ such that the graph is also endowed with a weighted adjacency matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$ and a Laplacian matrix $\mathbf{L} = \mathrm{Diag}(\mathbf{A}\mathbf{1}) - \mathbf{A}$. Let $d_{\max}$ denote the maximum degree in $\mathcal{G}$. The Laplacian matrix admits the eigendecomposition $\mathbf{L} = \mathbf{V}\Lambda\mathbf{V}^\top$ such that $\Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_N)$ has eigenvalues sorted in ascending order $0 = \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$.

This work focuses on analyzing the graph topology learning using a strategy similar to [1], [2]. In this framework, we assume that the graph signals $\mathbf{y}_1, \ldots, \mathbf{y}_M$ are smooth w.r.t. $\mathbf{L}$ such that $\mathbf{y}_m^\top \mathbf{L} \mathbf{y}_m \approx 0$ for any $m = 1, \ldots, M$. It inspires the graph learning problem:

$$\min_{\hat{\mathbf{L}}} J_f(\hat{\mathbf{L}}) := \textstyle\sum_{m=1}^M \mathbf{y}_m^\top \hat{\mathbf{L}} \mathbf{y}_m \quad \text{s.t. } \hat{\mathbf{L}} \in \mathcal{L}_N, \tag{1}$$

where

$$\mathcal{L}_N := \{\hat{\mathbf{L}} \in \mathbb{R}^{N \times N} : \mathrm{Tr}(\hat{\mathbf{L}}) = N, \hat{\mathbf{L}}\mathbf{1} = \mathbf{0}, \hat{\mathbf{L}} = \hat{\mathbf{L}}^\top\}.$$

We denote $\mathbf{L}^\star$ as an optimal solution to (1).

However, the above requires *full observations* of graph signals. To distinguish the above from the partial observation

case of interest, we let $\mathbf{y}_{o,1}, \ldots, \mathbf{y}_{o,M}$ be the set of partially observed graph signals such that

$$\mathbf{y}_{o,m} = \mathbf{E}_o \mathbf{y}_m, \text{ with } \mathbf{E}_o \in \mathbb{R}^{n \times N}$$

and $\mathbf{E}_o$ takes a subset of n nodes from $\mathcal{V}$. We consider the following hidden-node agnostic graph learning problem:

$$\min_{\hat{\mathbf{L}}_p} J_p(\hat{\mathbf{L}}_p) := \sum_{m=1}^{M} \mathbf{y}_{o,m}^\top \hat{\mathbf{L}}_p \mathbf{y}_{o,m} \quad \text{s.t. } \hat{\mathbf{L}} \in \mathcal{L}_n. \tag{2}$$

Notice that the above problem is identical to (1) except for the use of partial graph signals and $\hat{\mathbf{L}}_p$ is an estimate for the Laplacian matrix corresponding to a subgraph of $\mathcal{G}$ with nodes selected in $\mathbf{E}_o$. We denote $\mathbf{L}_p^\star$ as an optimal solution to (2).

Analysis. The goal of this work is to analyze the relationship between the optimal solutions to (1), (2). Particularly, we wish to show that $\mathbf{L}_p^\star$ corresponds to a column/row sampled version of $\mathbf{L}^\star$. To this end, we define:

$$\widehat{\mathbf{L}} := \frac{N}{n} \mathbf{E}_o^\top \mathbf{L}_p^\star \mathbf{E}_o, \ \widetilde{\mathbf{L}}_p := \underbrace{\mathbf{E}_o \mathbf{L}^\star \mathbf{E}_o^\top}_{=: \mathbf{L}_{oo}^\star} + \mathrm{Diag}(\mathbf{L}_{oh}^\star \mathbf{1})$$

where $\widehat{\mathbf{L}} \in \mathcal{L}_N$, $\widetilde{\mathbf{L}}_p \in \mathcal{L}_n$ and we defined $\mathbf{L}_{oh}^\star$ to be the submatrix of $\mathbf{L}^\star$ corresponding to edges between observable and hidden nodes. We further assume $\mathbf{L}_{oh}^\star \mathbf{1} \leq \epsilon \mathbf{1}$ for some small ϵ, and $\mathrm{Tr}(\mathbf{L}_{oo}^\star) - \mathbf{1}^\top \mathbf{L}_{oh}^\star \mathbf{1} \geq cn$ for constant $c > 0$.

As an illustration, our analysis concentrates on an ideal scenario where the (fully observed) graph signals $\mathbf{y}_m$ lie in a K-dimensional subspace, where $K \ll n \leq N$, with

$$\mathbf{y}_m \in \mathrm{span}(\mathbf{V}_K), \ m = 1, \ldots, M, \tag{3}$$

such that $\mathbf{V}_K$ that consists of K eigenvectors of $\mathbf{L}^\star$ with the K smallest eigenvalues. We note that the scenario is satisfied when solving (1) learns the ground truth graph topology. This is a plausible scenario for graph signals that are resulted from a low pass graph filtering process with cutoff bandwidth K [7]. Our main result under (3) is summarized as follows:

Theorem 1. *Consider a random partial observation set $\mathcal{S} = \{s(1), s(2), ..., s(n)\}$, which is sampled independently with replacements from the node index set $\{1, ..., N\}$. Then, with any $\delta \in (0, 1)$, there is an $t \in (0, 1)$ such that with probability at least $1-\delta$, provided that the number of observations satisfies*

$$\frac{n}{N} \geq \frac{3}{t^2} \max_{1 \leq i \leq N} \|\mathbf{V}_K^\top \mathbf{e}_i\|_2^2 \ln\left(\frac{2K}{\delta}\right),$$

the following inequalities hold

$$J_{\mathrm{p}}(\mathbf{L}_{\mathrm{p}}^\star) \leq J_{\mathrm{p}}(\widetilde{\mathbf{L}_{\mathrm{p}}}) \leq \frac{1+t}{c} \frac{\sigma_{\max}(\mathbf{L})}{\sigma_{\min}^+(\mathbf{L})} J_{\mathrm{p}}(\mathbf{L}_{\mathrm{p}}^\star) + \mathcal{O}\left(\frac{\epsilon}{c}\right)$$

and

$$J_{\mathrm{f}}(\mathbf{L}^\star) \leq J_{\mathrm{f}}(\widehat{\mathbf{L}}) \leq \frac{1+t}{c} \frac{\sigma_{\max}(\mathbf{L})}{\sigma_{\min}^+(\mathbf{L})} J_{\mathrm{f}}(\mathbf{L}^\star) - \mathcal{O}\left(\frac{N\epsilon}{cn}\right).$$

We observe that if $\frac{1+t}{c} \frac{\sigma_{\max}(\mathbf{L})}{\sigma_{\min}^+(\mathbf{L})} = \Theta(1)$, then the above theorem suggests that $\mathbf{L}_p^\star$ corresponds to a row/column sampled version of $\mathbf{L}^\star$. In this case, such condition can be satisfied for non-modularized graphs. Importantly, our result is insensitive to the number of hidden nodes, suggesting that the conclusion continues to hold even when $N - n \gg 1$. To give insights, our theorem is achieved by establishing the following one-sided RIP property, i.e., with high probability and for any $\mathbf{y} \in \mathrm{span}(\mathbf{V}_K)$,

$$\mathbf{y} \mathbf{E}_o^\top \mathbf{E}_o \mathbf{L} \mathbf{E}_o^\top \mathbf{E}_o \mathbf{y} \leq (1+t) \frac{n}{N} \frac{\sigma_{\max}(\mathbf{L})}{\sigma_{\min}^+(\mathbf{L})} \mathbf{y}^\top \mathbf{L} \mathbf{y}. \tag{4}$$

A proof sketch can be found in the appendix.

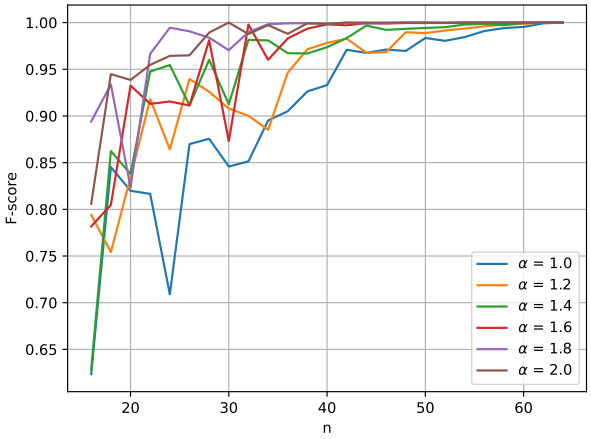

Fig. 1. F-score between $\mathbf{L}_p^\star$ learned from partial signals and the observed part of $\mathbf{L}^\star$, varying with number of observed nodes n

Preliminary Numerical Results. Lastly, we validate our theoretical findings by considering a numerical toy example with synthetically generated graph signals data. We let $\mathcal{G}$ be an Erdos-Renyi graph with connectivity $p = 0.3$, $N = 64$ nodes, and $\mathbf{y}_m$ be a low-pass graph signals generated as $\mathbf{y}_m = (\mathbf{I} + \alpha \mathbf{L})^{-1} \mathbf{x}_m$, where $\alpha > 0$ controls the low pass-ness of the graph filter; the higher α is, the more low pass the resulting signals are. Increasing α implies a higher chance for (3) to be satisfied. We consider the graph learning problem $M = 20$ samples. Fig. 1 shows the median F-scores of $\mathbf{L}_p^\star$ compared with the corresponding observed part of $\mathbf{L}^\star$ against n. Observe that even when $n \ll N$, the graph learnt from (2) remains accurate compared to the benchmark with (1), thus validating Theorem 1. We can also see that the more low-pass the graph signals are, a good F-score is achieved with a smaller number of observations n.

III. Conclusion & Future Work

We have presented the preliminary theoretical result that demonstrates the robustness of graph topology learning based on the smoothness criterion (2) against partial observation. We anticipate that the results can be extended to other tasks related to graph topology learning such as community detection [8], [9] and the bound can be further tightened. The above extensions are part of the ongoing work which we expect to deliver during the workshop.

REFERENCES

[1] X. Dong, D. Thanou, P. Frossard, and P. Vandergheynst, "Learning laplacian matrix in smooth graph signal representations," *IEEE Transactions on Signal Processing*, vol. 64, no. 23, pp. 6160–6173, 2016.

[2] V. Kalofolias, "How to learn a graph from smooth signals," in *Artificial intelligence and statistics*. PMLR, 2016, pp. 920–929.

[3] X. Dong, D. Thanou, M. Rabbat, and P. Frossard, "Learning graphs from data: A signal representation perspective," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 44–63, 2019.

[4] G. Mateos, S. Segarra, A. G. Marques, and A. Ribeiro, "Connecting the dots: Identifying network structure via graph signal processing," *IEEE Signal Processing Magazine*, vol. 36, no. 3, pp. 16–43, 2019.

[5] V. Chandrasekaran, P. A. Parrilo, and A. S. Willsky, "Latent variable graphical model selection via convex optimization," *The Annals of Statistics*, pp. 1935–1967, 2012.

[6] A. Buciulea, S. Rey, and A. G. Marques, "Learning graphs from smooth and graph-stationary signals with hidden variables," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 8, pp. 273–287, 2022.

[7] R. Ramakrishna, H.-T. Wai, and A. Scaglione, "A user guide to low-pass graph signal processing and its applications: Tools and applications," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 74–85, 2020.

[8] H.-T. Wai, Y. C. Eldar, A. E. Ozdaglar, and A. Scaglione, "Community inference from partially observed graph signals: Algorithms and analysis," *IEEE Transactions on Signal Processing*, vol. 70, pp. 2136–2151, 2022.

[9] H.-S. Nguyen and H.-T. Wai, "On detecting low-pass graph signals under partial observations," in *2024 IEEE 13rd Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2024, pp. 1–5.

[10] J. A. Tropp, "User-friendly tail bounds for sums of random matrices," *Foundations of Computational Mathematics*, vol. 12, no. 4, p. 389–434, 2011.

APPENDIX: OMITTED PROOFS

Lemma 1. *Consider a random partial observation $\mathcal{S} = \{s(1), s(2), ..., s(n)\}$, which is sampled independently with replacements from the node indices $\{1, ..., N\}$. The set $\mathcal{S}$ is encoded in a matrix $\mathbf{E}_o \in \{0,1\}^{n\times N}$ that gives $\mathbf{L}_{oo} = \mathbf{E}_o\mathbf{L}\mathbf{E}_o^\top \notin \mathcal{L}_n$ and $\mathbf{y}_o = \mathbf{E}_o\mathbf{y} \in \mathbb{R}^n$. For any $\delta \in (0,1)$, there exists a $t \in (0,1)$ such that with probability at least $1-\delta$,*

$$\mathbf{y}_o^\top \mathbf{L}_{oo}\mathbf{y}_o \leq (1+t)\frac{n}{N}\frac{\sigma_{\max}(\mathbf{L})}{\sigma^+_{\min}(\mathbf{L})}\mathbf{y}^\top\mathbf{L}\mathbf{y},\ \forall \mathbf{y} \in \text{span}(\mathbf{V}_K), \tag{5}$$

provided that

$$\frac{n}{N} \geq \frac{3}{t^2}\max_{1\leq i\leq N}\|\mathbf{V}_K^\top\mathbf{e}_i\|_2^2 \ln\left(\frac{2K}{\delta}\right).$$

Proof. To begin, we notice that

$$\begin{aligned}\mathbf{y}_o^\top \mathbf{L}_{oo}\mathbf{y}_o &= \mathbf{y}^\top\mathbf{E}_o^\top\mathbf{L}_{oo}\mathbf{E}_o\mathbf{y}\\ &\leq \|\mathbf{L}_{oo}\|_2(\mathbf{y}^\top\mathbf{E}_o^\top\mathbf{E}_o\mathbf{y})\\ &\leq \|\mathbf{L}\|_2(\mathbf{y}^\top\mathbf{E}_o^\top\mathbf{E}_o\mathbf{y}).\end{aligned}$$

Let $\boldsymbol{X}_\ell := \frac{N}{n}\mathbf{V}_K^\top\mathbf{e}_{s(\ell)}\mathbf{e}_{s(\ell)}^\top\mathbf{V}_K$, whose sum is

$$\begin{aligned}\boldsymbol{X} := \frac{N}{n}\sum_{\ell=1}^n \boldsymbol{X}_\ell &= \frac{N}{n}\mathbf{V}_K^\top\left(\sum_{\ell=1}^n \mathbf{e}_{s(\ell)}\mathbf{e}_{s(\ell)}^\top\right)\mathbf{V}_K\\ &= \frac{N}{n}\mathbf{V}_K^\top\mathbf{E}_o^\top\mathbf{E}_o\mathbf{V}_K.\end{aligned}$$

Note the followings:

$$\begin{aligned}&\mathbb{E}[\boldsymbol{X}_\ell] = \tfrac{N}{n}\mathbf{V}_K^\top\mathbb{E}[\mathbf{e}_{s(\ell)}\mathbf{e}_{s(\ell)}^\top]\mathbf{V}_K = \tfrac{1}{n}\mathbf{V}_K^\top\mathbf{V}_K = \tfrac{1}{n}\mathbf{I},\\ &\mu_{\max} := \lambda_{\max}\left(\textstyle\sum_{\ell=1}^n \mathbb{E}[\boldsymbol{X}_\ell]\right) = 1,\\ &\lambda_{\max}(\boldsymbol{X}_\ell) = \tfrac{N}{n}\lambda_{\max}(\mathbf{V}_K^\top\mathbf{e}_{s(\ell)}\mathbf{e}_{s(\ell)}^\top\mathbf{V}_K) \leq \tfrac{N}{n}\max_i\|\mathbf{V}_K^\top\mathbf{e}_i\|_2^2.\end{aligned}$$

Then, the matrix Chernoff's bound [10, Corollary 5.2] states that for any $\delta \in (0,1)$, with probability at least $1-\delta$,

$$\|\boldsymbol{X}\|_2 \leq 1 + t, \tag{6}$$

provided that

$$\frac{n}{N} \geq \frac{2}{t^2}\max_{1\leq i\leq N}\|\mathbf{V}_K^\top\mathbf{e}_i\|_2^2 \ln\left(\frac{2K}{\delta}\right).$$

Observe that (6) implies

$$\frac{N}{n}\|\mathbf{E}_o\mathbf{V}_K\boldsymbol{v}\|_2^2 \leq (1+\delta)\|\boldsymbol{v}\|_2^2,$$

which means for any $\mathbf{y} \in \text{span}(\mathbf{V}_K)$,

$$\frac{N}{n}\|\mathbf{E}_o\mathbf{y}\|_2^2 \leq (1+\delta)\|\mathbf{y}\|_2^2,$$

Therefore, with high probability, $\mathbf{y}_o^\top\mathbf{L}_{oo}\mathbf{y}_o$ can be bounded as

$$\begin{aligned}\mathbf{y}_o^\top\mathbf{L}_{oo}\mathbf{y}_o &\leq (1+\delta)\|\mathbf{L}\|_2\|\mathbf{y}\|_2^2\\ &\leq (1+\delta)\|\mathbf{L}\|_2\|(\mathbf{L}^{1/2})^\dagger\|_2^2\|\mathbf{L}^{1/2}\mathbf{y}\|_2^2\\ &= (1+\delta)\frac{n}{N}\frac{\sigma_{\max}(\mathbf{L})}{\sigma^+_{\min}(\mathbf{L})}\mathbf{y}^\top\mathbf{L}\mathbf{y}.\end{aligned}$$

□

Proof of Theorem 1. Let us define some functions and terms:

$$\begin{aligned}&J_{\rm f}(\mathbf{L}) := \tfrac{1}{m}\textstyle\sum_{i=1}^m \mathbf{y}_i^\top\mathbf{L}\boldsymbol{y}_i, \quad J_{\rm p}(\mathbf{L}_{\mathsf{p}}) := \tfrac{1}{m}\textstyle\sum_{i=1}^m \mathbf{y}_{o,i}^\top\mathbf{L}_{\mathsf{p}}\boldsymbol{y}_{o,i},\\ &\mathbf{L}^\star := \arg\min_{\mathbf{L}\in\mathcal{L}_N} J_{\rm f}(\mathbf{L}), \quad \mathbf{L}_{\mathsf{p}}^\star = \arg\min_{\mathbf{L}_{\mathsf{p}}\in\mathcal{L}_n} J_{\rm p}(\mathbf{L}_{\mathsf{p}}).\end{aligned}$$

Furthermore, we note that $\widehat{\mathbf{L}} := \frac{N}{n}\mathbf{E}_o^\top\mathbf{L}_{\mathsf{p}}^\star\mathbf{E}_o \in \mathcal{L}_N$ and

$$\widetilde{\mathbf{L}_{\mathsf{p}}} = \frac{n}{\mathbf{1}^\top\mathbf{L}_{oh}^\star\mathbf{1} + \text{Tr}(\mathbf{L}_{oo}^\star)}\left(\mathbf{E}_o\mathbf{L}^\star\mathbf{E}_o^\top + \text{Diag}(\mathbf{L}_{oh}^\star\mathbf{1})\right) \in \mathcal{L}_n.$$

Denote $C(t) := (1+t)\frac{n}{N}\frac{\sigma_{\max}(\mathbf{L})}{\sigma^+_{\min}(\mathbf{L})}$. Applying Lemma 1 implies:

$$\begin{aligned}&J_{\rm f}(\mathbf{L}^\star) \leq J_{\rm f}(\widehat{\mathbf{L}})\\ &= \frac{1}{m}\sum_{i=1}^m \frac{N}{n}\mathbf{y}_{o,i}^\top\mathbf{L}_{\mathsf{p}}^\star\mathbf{y}_{o,i}\\ &= \frac{N}{n}J_{\rm p}(\mathbf{L}_{\mathsf{p}}^\star)\\ &\leq \frac{N}{n}J_{\rm p}(\widetilde{\mathbf{L}_{\mathsf{p}}})\\ &= \frac{N}{\mathbf{1}^\top\mathbf{L}_{oh}^\star\mathbf{1} + \text{Tr}(\mathbf{L}_{oo}^\star)}\left(J_{\rm p}(\mathbf{E}_o\mathbf{L}^\star\mathbf{E}_o^\top) + J_{\rm p}(\text{Diag}(\mathbf{L}_{oh}^\star\mathbf{1}))\right)\\ &\leq \frac{N}{\mathbf{1}^\top\mathbf{L}_{oh}^\star\mathbf{1} + \text{Tr}(\mathbf{L}_{oo}^\star)}\left(C(t)J_{\rm f}(\mathbf{L}^\star) + J_{\rm p}(\text{Diag}(\mathbf{L}_{oh}^\star\mathbf{1}))\right).\end{aligned}$$

Overall, this gives the approximation bound

$$J_{\rm f}(\mathbf{L}^\star) \leq J_{\rm f}(\widehat{\mathbf{L}}) \leq \frac{N}{\text{Tr}(\mathbf{L}_{oo}^\star) + \mathbf{1}^\top\mathbf{L}_{oh}^\star\mathbf{1}}\left(C(t)J_{\rm f}(\mathbf{L}^\star) - \mathcal{O}(\epsilon)\right).$$

Conversely, we can also bound that

$$
\begin{aligned}
&J_{\mathrm{p}}(\mathbf{L}_{\mathsf{p}}^{\star}) \leq J_{\mathrm{p}}(\widetilde{\mathbf{L}_{\mathsf{p}}}) \\
&= \frac{n}{\mathbf{1}^{\top}\mathbf{L}_{oh}^{\star}\mathbf{1} + \mathrm{Tr}(\mathbf{L}_{oo}^{\star})} \left(J_{\mathrm{p}}(\mathbf{E}_o \mathbf{L}^{\star} \mathbf{E}_o^{\top}) + J_{\mathrm{p}}(\mathrm{Diag}(\mathbf{L}_{oh}^{\star}\mathbf{1}))\right) \\
&\leq \frac{n}{\mathbf{1}^{\top}\mathbf{L}_{oh}^{\star}\mathbf{1} + \mathrm{Tr}(\mathbf{L}_{oo}^{\star})} \left(C(t) J_{\mathrm{f}}(\mathbf{L}^{\star}) + J_{\mathrm{p}}(\mathrm{Diag}(\mathbf{L}_{oh}^{\star}\mathbf{1}))\right) \\
&\leq \frac{n}{\mathbf{1}^{\top}\mathbf{L}_{oh}^{\star}\mathbf{1} + \mathrm{Tr}(\mathbf{L}_{oo}^{\star})} \left(C(t) J_{\mathrm{f}}(\widehat{\mathbf{L}}) + J_{\mathrm{p}}(\mathrm{Diag}(\mathbf{L}_{oh}^{\star}\mathbf{1}))\right) \\
&= \frac{n}{\mathbf{1}^{\top}\mathbf{L}_{oh}^{\star}\mathbf{1} + \mathrm{Tr}(\mathbf{L}_{oo}^{\star})} \left(C(t) \frac{N}{n} J_{\mathrm{p}}(\mathbf{L}_{\mathsf{p}}^{\star}) + J_{\mathrm{p}}(\mathrm{Diag}(\mathbf{L}_{oh}^{\star}\mathbf{1}))\right) \\
&= \frac{N}{\mathbf{1}^{\top}\mathbf{L}_{oh}^{\star}\mathbf{1} + \mathrm{Tr}(\mathbf{L}_{oo}^{\star})} C(t) J_{\mathrm{p}}(\mathbf{L}_{\mathsf{p}}^{\star}) \\
&+ \frac{n}{\mathbf{1}^{\top}\mathbf{L}_{oh}^{\star}\mathbf{1} + \mathrm{Tr}(\mathbf{L}_{oo}^{\star})} J_{\mathrm{p}}(\mathrm{Diag}(\mathbf{L}_{oh}^{\star}\mathbf{1})).
\end{aligned}
$$

Taking $\mathbf{L}_{oh}^{\star}\mathbf{1} = \mathcal{O}(\epsilon)$ yields the bound

$$
\begin{aligned}
&J_{\mathrm{p}}(\mathbf{L}_{\mathsf{p}}^{\star}) \leq J_{\mathrm{p}}(\widetilde{\mathbf{L}_{\mathsf{p}}}) \\
&\leq \frac{N}{\mathrm{Tr}(\mathbf{L}_{oo}^{\star}) + \mathbf{1}^{\top}\mathbf{L}_{oh}^{\star}\mathbf{1}} C(t) J_{\mathrm{p}}(\mathbf{L}_{\mathsf{p}}^{\star}) + \frac{n}{\mathrm{Tr}(\mathbf{L}_{oo}^{\star}) + \mathbf{1}^{\top}\mathbf{L}_{oh}^{\star}\mathbf{1}} \mathcal{O}(\epsilon).
\end{aligned}
$$

□