SMOOTHING COMPLEX-VALUED SIGNALS ON GRAPHS WITH MONTE-CARLO

Hugo Jaquard\textsuperscript{†‡}  Michael Fanuel\textsuperscript{†‡}  Pierre-Olivier Amblard\textsuperscript{†}  Rémi Bardenet\textsuperscript{‡}  Simon Barthelmé\textsuperscript{†}  Nicolas Tremblay\textsuperscript{†}

\textsuperscript{†} GIPSA-lab, CNRS, Univ. Grenoble Alpes, Grenoble INP  
\textsuperscript{‡} Univ. Lille, CNRS, Centrale Lille, UMR 9189 – CRISTAL

\textsuperscript{*} equally contributing authors

ABSTRACT

We introduce new smoothing estimators for complex signals on graphs, based on a recently studied Determinantal Point Process (DPP). These estimators are built from subsets of edges and nodes drawn according to this DPP, making up trees, unicycles, i.e., connected components containing exactly one cycle. We provide a Julia implementation of these estimators and study their performance when applied to a ranking problem.

Index Terms— magnetic laplacian, spanning forests, determinantal point processes, graph smoothing, angular synchronization, ranking

1. INTRODUCTION

Graph signal processing (GSP, [1]) usually considers real data defined over the nodes of a graph $G = (V,E)$, and classically relies on the graph Laplacian. For instance, a typical task in GSP consists in smoothing (denoising) a signal $g \in \mathbb{R}^V$ by solving the penalized (Tikhonov) problem

$$
\arg\min_{f \in \mathbb{R}^V} q\|f - g\|^2 + f^\top L f
$$

where $L \in \mathbb{R}^{V \times V}$ is the graph Laplacian. We will consider weighted and undirected graphs. In Equation (1), $f^\top L f$ penalises the squared-norm of the discrete derivative on the graph, i.e.:

$$
f^\top L f = \sum_{e=(v,v')} w_e (f(v') - f(v))^2,
$$

where the sum runs over all edges of the graph and $w_e > 0$ is the edge weight. Thus, the quadratic form $f^\top L f$ can be thought of as i/ computing the squared difference between signal values along neighbouring edges, ii/ computing a weighted sum.

For multivariate signals, defining a discrete derivative becomes less obvious, as additional geometry enters the picture. For instance, taking $f(v') - f(v)$ along an edge $e = (v,v')$ assumes that $f(v')$ and $f(v)$ use the same coordinate system, and introducing a local change of basis along the edge $(v,v')$ is a way to precise the relation between $f(v')$ and $f(v)$. We focus on complex-valued signals $f \in \mathbb{C}^V$, for which a multiplication by $e^{i\theta_e}$, where $\theta_e \in [0,2\pi]$, can for instance represent a known phase offset between measurements $f(v')$ and $f(v)$. In this setting, we can define a magnetic Laplacian $L \in \mathbb{C}^{V \times V}$ [3] which acts as follows ($f^*$ denotes the conjugate transpose of $f$):

$$
f^* L f = \sum_{e=(v,v')} w_e |f(v') - e^{i\theta_e} f(v)|^2.
$$

Note that this equation supposes an orientation of each edge: each $\theta_e$ is thus given with an orientation of $e$. This choice of orientation is however arbitrary: for a given edge $e$, one orientation associated to $\theta_e$ is equivalent to the other orientation associated to $-\theta_e$ as $|f(v') - e^{i\theta_e} f(v)|^2 = |e^{-i\theta_e} f(v') - f(v)|^2$. The set $\{e^{i\theta_e}\}_e$ describes a unitary connection between the nodes of $G$ [3].

Such Laplacians have applications to synchronization [4] and ranking problems [5–7], as we explain in Section 4.

GSP algorithms that use the graph Laplacian often have $O(|V|^3)$ scaling when implemented exactly, due to the matrix inversions or factorisations that are used (e.g., the solution of Equation (1) reads $f_o = q(L + qI)^{-1} g$). In large graphs approximate methods are necessary, and [8] introduced a Monte-Carlo estimator for the Tikhonov smoothing problem of Equation (1), with favourable asymptotic runtime. In this work we generalise the estimator of [8] to complex signals, using a process recently introduced in [9].

In Section 2, we introduce a slight variation of the random process of [9], defined over graphs with a unitary connection. Our main theoretical results are given in Section 3, where we derive estimators for the solution of the Tikhonov smoothing
problem. We describe a practical application to the ranking problem in Section 4. An extended version of this paper with technical definitions and detailed proofs is available on arXiv [10].

2. A PROCESS OVER MULTI-TYPE SPANNING FORESTS

We are interested in a distribution generalizing both the uniform distribution over spanning trees (UST) of a graph, and the random spanning forests distribution [11]. A spanning tree is a subset of edges \( \phi \subseteq E \) such that the graph with nodes \( V \) and edges \( \phi \) is both connected and without cycles (cycle-free). A rooted spanning forest (RSF) \( \phi \subseteq E \cup V \) is the combination of a spanning forest and a set of distinguished nodes called the roots (one root per tree). A variation of these distributions can be defined over the set of spanning forests of unicycles (FU) [3, 12], subsets of edges containing exactly one cycle per connected component, and spanning all the nodes in \( V \). The generalization we consider instead draws its samples from the set of rooted multi-type spanning forests (MTSF). A rooted MTSF is a spanning subset of edges and nodes \( \phi \subseteq E \cup V \) whose connected components are made up of rooted trees and unicycles. These different structures are illustrated in Fig. 1.

We will use the following distribution, introduced in [9], over rooted MTSFs \( \phi = \phi_\bullet \cup \rho(\phi) \) of \( G \):

\[
\mathbb{P}(\phi) \propto \prod_{r \in \rho(\phi)} q_r \prod_{v \in \phi_\bullet} w_v \prod_{e} (2 - 2 \cos(\theta_e)),
\]

for which \( 2 - 2 \cos(\theta_e) \) is large, meaning that these cycles are inconsistent. Part of the answer lies in the fact that Equation (3) is known to describe a Determinantal Point Process (DPP) over \( E \cup V \), a useful property for proving the results in Section 3.

Aforementioned distributions such as USTs and RSFs are conveniently sampled using (a variant of) Wilson’s algorithm [13], based on random walks on the graph (see Appendix B.2 of [10] for the pseudo-code). These sampling procedures have been generalized to FUs [14] and, recently, to MTSFs [9], under the sampling condition:

\[
\cos(\theta_e) \geq 0 \text{ for all cycles } \gamma.
\]

Note that this condition applies to all cycles \( \gamma \) in \( G \), and not only in some MTSF \( \phi \). Bounds on the running time of the sampling algorithm are discussed in [9, 14]. Here, we only mention that the expected running time is linear in the number of edges, and that it decreases as \( \min_v q_v \) increases.

3. ESTIMATORS FOR SMOOTHING

Monte-Carlo estimators for smoothing on graphs have been developed using RSFs [8]. The main idea is to sample a rooted RSF \( \phi \) before propagating the value of the function at the roots to the other nodes in their associated trees. In the following, we show how this can be generalized to graphs with a unitary connections using the distribution in Equation (3). Specifically, given \( q \in \mathbb{C}^V \) we derive estimators of \( f_0 = (L+Q)^{-1}Qg \). Here \( Q \) is the diagonal matrix of the \( q_v \)'s and \( L = D - A_\theta \) is the Hermitian magnetic Laplacian matrix, with \( D \) the diagonal degree matrix and \( (A_\theta)_{e', e} = e^{i\theta_e} \) if \( e = (v, v') \in E \) (0 if \( e \notin E \)). When \( q_e = q > 0 \) for all \( v \in V \), \( f_0 \) is the optimal solution of the Tikhonov problem:

\[
\argmin_{f \in \mathbb{C}^V} \|f - g\|^2 + \lambda^2 Lf.
\]

The first estimator \( \tilde{f} \) is built by propagating values through the transport maps \( z \mapsto e^{i\theta_e}z \) on a rooted multi-type spanning forest \( \phi \) sampled according to Equation (3):

\[
\tilde{f}(v) = \begin{cases} 
\psi_{r_\phi(v)} \varphi_{\phi_\bullet(v)}(\phi_\bullet(v)) & \text{if the connected component of } v \text{ is a rooted tree}, \\
0 & \text{otherwise},
\end{cases}
\]

where \( r_\phi : V \to V \) maps nodes to the root of the tree containing them, \( a \to b \) denotes the unique path from \( a \) to \( b \) in \( \phi \), and \( \psi_{a \to b} = \prod_{e \in a \to b} e^{i\theta_e} \).

Proposition 1. \( \tilde{f} \) is an unbiased estimator of \( f_0 \):

\[
\mathbb{E}_{\phi}(\tilde{f}) = f_0.
\]

As a consequence of the Central Limit Theorem, Monte-Carlo estimators converge at a \( O(\frac{1}{\sqrt{n}}) \) rate, with \( \sigma \) the standard deviation of the estimator and \( n \) the number of samples.
Instead of increasing the number of samples one may instead focus on decreasing the variance, and in our setting this can be done at little additional cost. A first approach is to use a Rao-Blackwell version of the estimator \( \hat{f} \) (see [15, 16]) by conditioning on the set of unrooted connected components \( \pi \subseteq E \) of the MTSF. Consider the estimator:

\[
\tilde{f}(v) = \begin{cases} 
\lambda(\psi_{\psi,v}(v), r_{\psi}(v)) & \text{if the connected component of } v \text{ is a rooted tree,} \\
0 & \text{otherwise,}
\end{cases}
\]

where \( h(u) = \sum_{w \in T_v} q_u \psi_{u \rightarrow w} g(w) \), with \( T_v \) the tree containing \( v \) in \( \phi \). Variance reduction is achieved solely from computing a mean over the nodes of the rooted trees.

**Proposition 2.** We have \( \forall v \in V, \tilde{f}(v) = E_{\phi}(\hat{f}(v)|\phi_o = \pi) \) and, as a consequence: \( E_\pi(\tilde{f}) = E_{\phi}(\hat{f}) = f_o \). Also, by the law of total variance, \( \tilde{f}(v) \) has a lower variance than \( f(v) \):

\[
\text{Var}(\tilde{f}(v)) = \text{Var}(\hat{f}(v)) - E(\text{Var}(\hat{f}(v)|\phi_o = \pi)) \leq \text{Var}(\hat{f}(v))
\]

The method of control variates is another classical variance-reduction technique for Monte-Carlo estimators, adding a term with zero mean to obtain a modified estimator with the same expectation but lower variance [17]. The following is proved in [18], when \( q_v = q \) for all \( v \in V \) (i.e. \( Q = qI \)).

**Proposition 3.** Set \( \alpha = \frac{2q}{q + 2d_m} \), where \( d_m \) is the maximum degree in the graph. Then, the estimator

\[
\hat{f} = \tilde{f} - \alpha(q^{-1}(L + qI)\tilde{f} - g)
\]

is an unbiased estimator of \( f_o \) and verifies

\[
\forall v \in V \quad \text{Var}(\hat{f}(v)) \leq \text{Var}(\tilde{f}(v))
\]

For a graph with heterogeneous degree distribution, with e.g. a maximum degree much larger than the mean degree, \( \hat{f} \) is only a marginal improvement over \( \tilde{f} \). However, on graphs with a nearly-homogeneous degree distribution, which is the case for the graphs considered in Section 4, we obtain substantial improvements over \( \tilde{f} \).

### 4. Ranking from Corrupted Measurements

In order to study the actual performance of our estimators, we describe an application to the ranking problem. We focus on the ordinal ranking problem, which asks to linearly order a set \( X \) of \( n \) elements according to an incomplete, possibly incoherent, set of pairwise ordinal comparisons \( C_{i,j} \in \{-1, 1\} \), for \( j > i \). From this data, build a graph \( G \) with \( n \) nodes, and a directed edge from \( i \) to \( j \) if \( C_{i,j} = 1 \) (resp. from \( j \) to \( i \) if \( C_{i,j} = -1 \)). Ranking according to [6, 7] then suggests to define the unitary connection \( \theta_{i,j} = \frac{\pi C_{i,j}}{n} \) with \( \delta \in (0, 1) \), and to perform **angular synchronization** [4, 19] by solving:

\[
\min_{(\omega_j)_j \in \mathbb{V}} \sum_{(v,v') \in E} |e^{j\omega_{v'}} - e^{j\omega_v}|^2.
\]

The optimal arguments \( \omega_j \) then describe an embedding of the \( n \) points onto the unit circle, from which we can extract a ranking (see [7] for more details).

In practice, solving such a non-convex optimization problem is difficult, and a spectral relaxation is considered instead:

\[
\min_{f \in C^n, \|f\|^2 = n} f^* \tilde{L} f,
\]

for \( \tilde{L} = D^{-1/2}L D^{-1/2} \) the normalized graph Laplacian, with \( D \) the diagonal degree matrix. The solution of this classical problem is the eigenvector associated to the smallest eigenvalue of \( \tilde{L} \), which can for instance be computed by iterating the map \( x \mapsto M x / \|M x\| \) for \( M = q(L + qI)^{-1} \), this is the power method [20]. Computing \( x \mapsto M x \) can either be performed directly by solving a linear system, or using the estimators \( \hat{f}, \tilde{f} \) and \( f \).

If we set \( \delta = \frac{1}{4} \), the sampling condition in Equation (4) is satisfied and fast sampling can be achieved using the variation of Wilson’s algorithm described in [9].

#### 4.1. Experimental results

We illustrate in Fig. 2 some results regarding the performance of this approach obtained with our Julia implementation\(^1\). We work with comparisons \( C_{i,j} \) randomly obtained from a ground-truth ranking \( r_{\text{GT}} \) according to the Erdős-Rényi Outliers model [7]: a comparison is observed with probability \( s \); if a comparison is observed, it follows the ranking \( r_{\text{GT}} \) with probability \( p \), or is chosen uniformly in \( \{-1, 1\} \) otherwise.

We focus on the computationally challenging dense case and set \( s = 0.8 \). Unless otherwise specified, we use \( q = 0.1 \).

The performance of the estimator \( \hat{f} \) in recovering the underlying ranking, using \( m = 5 \) MTSFs for each of the \( k = 10 \) iterations of the power method, is illustrated in Figs. 2a and 2b for two graphs of size \( n = 300 \) and \( p \in \{0.6, 0.9\} \). The eigenvector of \( L \) computed from the power method without using the estimator is also plotted. Here, the initialization vector \( y_0 \) used for the power method is a random embedding of the \( n \) points in the unit circle, spaced out with angle \( \frac{2\pi}{2n} \).

Runtime benchmarks are available in Figs. 2c and 2d for the smoothing problem (computation of \( M y_0 \) solved either using \( f \) or using a Cholesky decomposition, and for the computation of the eigenvector of \( M \) for the power method iteration using \( f \) or a Chebyshev decomposition, and for the Lanczos method. We display the mean running time over 100 measurements on fixed graphs of size \( n \in \{10, 10^2, 10^3, 10^4\} \).

We plot Kendall’s \( \tau \) coefficients [21] (the larger the better) for the rankings recovered from the power method in Figs. 2e, 2f and 2g for \( n = 3000 \) and varying \( q \), averaging over 20 realisations.

Reconstruction errors \( \|My - e(f_o)\| \) across varying \( m \), averaged over 20 runs for each of the estimators \( e(f_o) \) of \( f_o \), are in Fig. 2h with \( n = 3000 \).

\(^1\)https://gricad-gt.montblanc.mp.ch/gemis-gtlab.univ-grenoble-alpes.fr/
4.2. Discussion

The results in Figs. 2a and 2b compare the performance of the reconstructions obtained by the power method with and without using \( \hat{f} \), on only one realization of the graph. For these noise regimes, the ground truth ranking \( \tau_{GT} \) can no longer be recovered (which would result in a straight diagonal line). In both cases, the recovered ranking is slightly more spread out when using the estimator instead of the exact power method.

We obtain faster runtime in the smoothing problem than a standard Cholesky decomposition, in graphs with more than \( n \sim 10^3 \) nodes (Fig. 2c). For the eigenvector computation in Fig. 2d, cross-over occurs at \( n \sim 10^4 \) against the power method, while we do not outperform a direct Lanczos iteration. Note that performance for the eigenvector computation may be increased by sampling only one set of \( m = 5 \) MTSFs, used for all iterations of the power method. Moreover, our current implementation is far from optimal.

As \( q \) decreases, the \( \tau \) coefficient increases until it reaches a fixed value in Figs. 2e, 2f and 2g. Comparing of Figs. 2e and 2f shows that this behavior actually reflects the convergence of the power method, as the spectral gap of \( M \) is lower for larger \( q \), requiring more iterations to converge. This suggests that there is a trade-off for the best choice of \( q \): it should be as large as possible in order to reduce sampling time, and not too large so that it allows fast convergence of the power method. Note also that Kendall’s \( \tau \) seems to reach the value of \( p \) when convergence occurs.

Regarding the convergence of the estimators, all three versions of the estimators in Fig. 2h have linear decay in log-log-scale, which is characteristic of the \( O(1/\sqrt{n}) \) convergence rate of Monte-Carlo estimators. In our simulations, the estimator \( \hat{f} \) improves on the regular estimator by a factor of 10.

Throughout these experiments, we sampled \( m = 5 \) MTSFs to compute the estimators. Here, this choice was arbitrary, but we notice nonetheless that this is of the order of the \( O(\log |V|) \) uniform spanning trees necessary to obtain spectral sparsifiers for connection-free graphs [22], recently adapted to MTSFs in [9].

5. CONCLUSION

We define new estimators built by propagating values along edges sampled according to a recently introduced DPP, for the smoothing problem on graphs endowed with a unitary connection, thus generalizing previous existing approaches on graphs. The evaluation of these estimators on the ranking problem, using our Julia implementation, show that they can be advantageous for smoothing starting from moderately sized graphs, as compared to a Cholesky decomposition, while they do not improve on a Lanczos iteration for eigenvector computation. Nonetheless, the proposed estimators exhibit a potentially useful computational property that is uncommon among deterministic algorithms: their computation can be carried out without any prior global knowledge of the graph, from local neighbor queries only, which in some scenarios may be the only practical interaction. The choice of the parameters \( q \) and \( m \) is application-dependent, and requires further investigation. Possible extensions of this work include designing new applications of the proposed estimators, as well as generalizing this approach to higher-dimensional signals, where current arguments do not carry over.
6. REFERENCES

[1] David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst, “The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains,” IEEE signal processing magazine, vol. 30, no. 3, pp. 83–98, 2013.

[2] Amit Singer and H-T Wu, “Vector diffusion maps and the connection laplacian,” Communications on pure and applied mathematics, vol. 65, no. 8, pp. 1067–1144, 2012.

[3] Richard Kenyon, “Spanning forests and the vector bundle Laplacian,” The Annals of Probability, vol. 39, no. 5, pp. 1983–2017, 2011.

[4] Amit Singer, “Angular synchronization by eigenvectors and semidefinite programming,” Applied and computational harmonic analysis, vol. 30, no. 1, pp. 20–36, 2011.

[5] X Yu Stella, “Angular embedding: from jarring intensity differences to perceived luminance,” in 2009 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2009, pp. 2302–2309.

[6] Stella Yu, “Angular embedding: A robust quadratic criterion,” IEEE transactions on pattern analysis and machine intelligence, vol. 34, no. 1, pp. 158–173, 2011.

[7] Mihai Cucuringu, “Sync-rank: Robust ranking, constrained ranking and rank aggregation via eigenvector and sdp synchronization,” IEEE Transactions on Network Science and Engineering, vol. 3, no. 1, pp. 58–79, 2016.

[8] Yusuf Yiğit Pilavcı, Pierre-Olivier Amblard, Simon Barthelmé, and Nicolas Tremblay, “Graph Tikhonov regularization and interpolation via random spanning forests,” IEEE transactions on Signal and Information Processing over Networks, vol. 7, pp. 359–374, 2021.

[9] Michaël Fanuel and Rémi Bardenet, “Sparsification of the regularized magnetic Laplacian with multi-type spanning forests,” arXiv preprint arXiv:2208.14797, 2022.

[10] Hugo Jaquard, Michaël Fanuel, Pierre-Olivier Amblard, Rémi Bardenet, Simon Barthelmé, and Nicolas Tremblay, “Smoothing complex-valued signals on graphs with monte-carlo,” arXiv preprint arXiv:2210.08014, 2022.

[11] Luca Avena and Alexandre Gaudilliére, “On some random forests with determinantal roots,” Preprint: Weierstraß-Institut für Angewandte Analysis und Stochastik, vol. 1881, 2013.

[12] Robin Forman, “Determinants of Laplacians on graphs,” Topology, vol. 32, no. 1, pp. 35–46, 1993.

[13] David Bruce Wilson, “Generating random spanning trees more quickly than the cover time,” in Proceedings of the twenty-eighth annual ACM symposium on Theory of computing, 1996, pp. 296–303.

[14] Adrien Kassel and Richard Kenyon, “Random curves on surfaces induced from the Laplacian determinant,” The Annals of Probability, vol. 45, no. 2, pp. 932–964, 2017.

[15] David Blackwell, “Conditional expectation and unbiased sequential estimation,” The Annals of Mathematical Statistics, pp. 105–110, 1947.

[16] C Radhakrishna Rao, “Information and the accuracy attainable in the estimation of statistical parameters,” in Breakthroughs in statistics, pp. 235–247. Springer, 1992.

[17] Zdravko Botev and Ad Ridder, “Variance reduction,” Wiley statsRef: Statistics reference online, pp. 1–6, 2017.

[18] Yusuf Yiğit Pilavcı, Pierre-Olivier Amblard, Simon Barthelmé, and Nicolas Tremblay, “Variance reduction in stochastic methods for large-scale regularised least-squares problems,” in Proceedings of the 30th European Signal Processing Conference (EUSIPCO), 2022, pp. 1771–1775.

[19] Afonso S Bandeira, “Ten lectures and forty-two open problems in the mathematics of data science,” Lecture Notes, 2015.

[20] Yousef Saad, Numerical methods for large eigenvalue problems: revised edition, SIAM, 2011.

[21] Maurice G Kendall, “A new measure of rank correlation,” Biometrika, vol. 30, no. 1/2, pp. 81–93, 1938.

[22] Tali Kaufman, Rasmus Kyng, and Federico Solód, “Scalar and matrix Chernoff bounds from $l_{\infty}$-independence,” in Proceedings of the 2022 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA). SIAM, 2022, pp. 3732–3753.