A Time-Vertex Signal Processing Framework

Scalable processing and meaningful representations for time-series on graphs

Francesco Grassi¹, Andreas Loukas², Nathanaël Perraudin², and Benjamin Ricaud²

¹Politecnico di Torino, Torino, Italy
²Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Abstract—An emerging way to deal with high-dimensional non-euclidean data is to assume that the underlying structure can be captured by a graph. Recently, ideas have begun to emerge related to the analysis of time-varying graph signals. This work aims to elevate the notion of joint harmonic analysis to a full-fledged framework denoted as Time-Vertex Signal Processing, that links together the time-domain signal processing techniques with the new tools of graph signal processing. This entails three main contributions: (a) We provide a formal motivation for harmonic time-vertex analysis as an analysis tool for the state evolution of simple Partial Differential Equations on graphs. (b) We improve the accuracy of joint filtering operators by up-to two orders of magnitude. (c) Using our joint filters, we construct time-vertex dictionaries analyzing the different scales and the local time-frequency content of a signal. The utility of our tools is illustrated in numerous applications and datasets, such as dynamic mesh denoising and classification, still-video inpainting, and source localization in seismic events. Our results suggest that joint analysis of time-vertex signals can bring benefits to regression and learning.

Index Terms—Time-Vertex Signal Processing, Graph Signal Processing, Partial Differential Equations

I. INTRODUCTION

Whether examining consensus and rumor spreading over social networks [1], transportation networks [2] and related epidemic spreading [3], neuronal activation patterns in the brain [4], as well as other datasets collected from a variety of fields, such as physics, engineering, and life-science, much of the high-dimensional data exhibit complex non-euclidean properties.

An emerging way to deal with these issues is to use a graph to capture the structure underlying the data. This has been the driving force behind recent efforts in the signal processing field to extend harmonic analysis to graph signals, i.e., signals supported on the vertices of irregular graphs [5], [6]. In the field of graph signal processing (GSP), the introduction of the graph Fourier transform (GFT) has enabled us to perform harmonic analysis taking into account the structure of the data, and has lead to improvements for tasks such as clustering [7], low-rank extraction [8], spectral estimation [9], [10], non-stationary analysis [11], [12] and semi-supervised learning [13], [14].

Nevertheless, though state-of-the-art graph-based methods have been successful for many tasks, so far they predominantly ignore the time-dimension of data, for example by treating successive signals independently or performing a global average [4], [9], [15]. On the contrary, many of the systems to which GSP is applied to are inherently dynamic.

Recently, several ideas begin to emerge related to the analysis of time-varying graph signals, such as Joint time-vertex Fourier transform (JFT) [16] and the joint time-vertex filters [17] and filterbanks [18]. While these constitute notable contributions, we argue that the potential of joint harmonic analysis is yet unexplored, both in terms of its foundations, algorithms, and applications.

In this work we aim at elevating the notion of joint harmonic analysis to a full-fledged framework, referred to as the Time-Vertex Signal Processing Framework, that links together the time-domain signal processing techniques with the new tools of GSP.

This entails the following contributions:

1. Connection to PDEs. We illustrate how joint analysis emerges when analyzing the state evolution of simple Partial Differential Equations (PDEs) on graphs (Section III-A). We also provide an example (epidemic spreading) demonstrating that the joint frequency analysis can be meaningful for the study non-linear and stochastic processes leading to a compact and intuitive representation (Section III-B).

2. Accurate fast joint filtering. In Section IV we illustrate the utility of joint filtering time-vertex signals and propose a fast filtering implementation, called Fast Fourier-Chebyshev (FFC) algorithm, which improves upon the state-of-the-art filters both in terms of separable and non-separable filtering objectives. For the latter case especially, our numerical experiments show that FFC can yield up to two orders of magnitude smaller error.

3. Overcomplete representations. We study redundant time-vertex dictionaries and exploit them for signal analysis and synthesis. The proposed framework includes a frame condition guaranteeing that no information is lost. Two particular cases are: time-vertex wavelets capturing the different scales of the signal components, and the
**short time-vertex Fourier transform** that is useful in determining the local time-frequency content of the signal.

4. **Illustrating the utility of time-vertex analysis.** Finally, Section VI provides experimental evidence for the utility of joint harmonic analysis in a number of graph-temporal datasets that were up to now not fully exploited, such as dynamic meshes, video and general dynamics over networks. The range of applications covers the classical signal processing problems of denoising, inpainting and compression, but also extends to feature extraction for classification and source localization problems.

### A. Related Work

The time-vertex framework is intimately linked with the stochastic analysis of multivariate signals and, therefore, with graphical models (e.g. [19] and references therein). The main difference between graphical models and GSP lies in the assumption about the relation between the signal and graph [20]. Graphical models adopt a purely Bayesian setting, where edges denote conditional dependencies between variables. As such, the graph usually is a proxy for the covariance and is learned from the data. On the other hand, GSP assumes that the graph is given and its relation to the signal can be understood through harmonic analysis.

In this context, the idea of time-vertex analysis can be traced back to the study [21] aiming to process multi-modal signals with different graphs associated with each of their modalities (i.e., one can consider a time-vertex signal as multi-modal, with time and graph being the two modalities). Collaboration between the graph theory and signal processing communities has led to new tools to process and analyze time-varying graphs and signals on a graph, such as multilayer graphs and tensor products of graphs [22]–[24]. The notion of joint harmonic analysis in a number of graph-temporal datasets that were up to now not fully exploited, such as dynamic meshes, video and general dynamics over networks. The range of applications covers the classical signal processing problems of denoising, inpainting and compression, but also extends to feature extraction for classification and source localization problems.

**Visualization, filtering and stationarity.** The idea of analyzing the behavior of graph filters with time-varying signals first appeared in [25], showing that they could be analyzed by applying jointly a GFT and a Z-transform and as such they possess a joint frequency response. Since then, we have seen a number of works dealing with time-varying signals on graphs: Authors in [26] propose a method that relies on graph wavelet theory and product graphs to visualize time-varying data defined on the vertices of a graph in order to identify spatial and/or temporal variations. A step towards the graphical model has been carried out by authors in [27]. In this work, authors assume data time dependencies to be modeled by an auto-regressive (AR) process and they propose several algorithms to estimate the network structure capturing the spatio-temporal dependencies and the coefficients of the AR process expressed as graph polynomial filters. In order to deal with the high computational complexity of the eigendecomposition, different filtering approximation algorithms have been proposed, mainly based on polynomials: centralized and distributed joint filter 2D Chebychev polynomial [16], separable rational [17] implementations, and autoregressive models [28].

Finally, in parallel with this work, the authors extended the notions of time stationarity and the recent graph stationarity [9] to the joint time-vertex domain [29] providing a framework for the statistical signal processing of time-vertex signals. Authors showed that assuming joint stationarity to regularize learning can yield significant accuracy improvements and reduce computational complexity in both estimation and recovery tasks with respect to purely time or graph methods [30], [31]. Despite the relevance of this work to time-vertex analysis, here we focus on the purely deterministic setting.

### II. HARMONIC TIME-VERTEX ANALYSIS

We denote by \( G = (V, E, W_G) \) the undirected graph, where \( V \) indicates the set of nodes, \( E \) the set of edges and \( W_G \) is the associated \( N \times N \) symmetric weight matrix. Furthermore, let \( L_G = D_G - W_G \) be the combinatorial Laplacian matrix, i.e. the finite difference approximation to the continuous Laplacian operator [14] or the Laplace-Beltrami operator for Riemannian manifolds [32]. We suppose that the signal on a graph is sampled at \( T \) successive regular intervals of unit length. If we denote by \( x_t \in \mathbb{R}^N \) the graph signal at instant \( t \), the time-varying graph signal corresponds to the matrix \( X = [x_1, x_2, \ldots, x_T] \in \mathbb{R}^{N \times T} \). We denote \( X^T, \ X, \) and \( X^* \) the transpose, the complex conjugate and the hermitian of \( X \). Furthermore, we refer to both \( X \) and its vectorized form \( x = \text{vec}(X) \in \mathbb{R}^{NT} \) as “time-vertex signal”.

#### A. The joint time-vertex Fourier transform

The main idea of harmonic analysis is to decompose a signal into oscillating modes thanks to the Fourier transform. For instance, one analyses oscillations along the temporal axis by applying the Discrete Fourier Transform (DFT) independently to each row of \( X \)

\[
\text{DFT}(X) = XU_T, \tag{1}
\]

where \( U_T \) is the normalized DFT matrix defined as

\[
U_T(t, k) = \frac{e^{-j \omega_k t}}{\sqrt{T}}, \quad \text{with} \quad \omega_k = \frac{2\pi(k-1)}{T}, \tag{2}
\]
Properties of JFT.

Property 1. JFT is an invertible transform. The inverse JFT in matrix and vector form are, respectively, \( J^{-1} \{ \mathbf{X} \} = \mathbf{U}_G \mathbf{X} \mathbf{U}_G^T \) and \( J^{-1} \{ \mathbf{x} \} = \mathbf{U}_G^T \mathbf{x} \).

Property 2. The Parseval relation holds:
\[
\sum_{n,t=1}^{N,T} |\mathbf{X}(n,t)|^2 = \sum_{\ell,k=1}^{N,T} |\tilde{\mathbf{X}}(\ell,k)|^2.
\]

Property 3. The transform is independent on the order GFT and DFT are applied to the time-vertex signal
\[
JFT\{ \mathbf{X} \} = GFT\{ DFT\{ \mathbf{X} \} \} = DFT\{ GFT\{ \mathbf{X} \} \}.
\]

Property 4. The subspace of zero graph and temporal frequency is spanned by the constant time-vertex signal \( \mathbf{1}^* \), with \( \mathbf{1} \) the all-ones vector.

with \( t, k = 1, 2, \ldots, T \). Similarly, the Graph Fourier Transform (GFT) [5], [11], [33] allows us to analyze oscillations along the graph edges. As each column of \( \mathbf{X} \) represents a time instant, the GFT of \( \mathbf{X} \) for all \( t \) reads
\[
GFT\{ \mathbf{X} \} = \tilde{\mathbf{X}} = \mathbf{U}_G^* \mathbf{X},
\]
where \( \mathbf{U}_G \) is obtained by the eigendecomposition \( \mathbf{L}_G = \mathbf{U}_G \mathbf{A}_G \mathbf{U}_G^T \) of the graph Laplacian. This spectral decomposition gives rise to a graph-specific notion of frequency as their squared modulus corresponds the Laplacian eigenvalue \( \mathbf{A}_G(\ell, \ell) = \lambda_\ell \).

Harmonic time-vertex analysis amounts to analyzing oscillations jointly along both the time and the vertex dimensions. Hence, assuming a non-varying graph in time, the joint time-vertex Fourier transform, or JFT for short, is obtained by applying the GFT on the graph dimension and the DFT along the time dimension [16].
\[
\tilde{\mathbf{X}} = JFT\{ \mathbf{X} \} = \mathbf{U}_G^* \mathbf{X} \mathbf{U}_T.
\]
Expressed in vector form, the transform becomes
\[
\hat{\mathbf{x}} = JFT\{ \mathbf{x} \} = \mathbf{U}_T^* \mathbf{x},
\]
where \( \mathbf{U}_J = \mathbf{U}_T \otimes \mathbf{U}_G \) is the Kronecker product of the basis. The relation between Eq.(4) and Eq.(5) is obtained through the property of the Kronecker product \( (\mathbf{M}_1 \otimes \mathbf{M}_2) \mathbf{x} = \mathbf{M}_2 \mathbf{X} \mathbf{M}_1^T \).

B. Time-vertex calculus and variation

In the following, we briefly present the main time-vertex differential operators. These will help us (a) to perform calculus on a finite, discrete time and space, and (b) to characterize the properties of the signals, such as smoothness, while taking into account the intrinsic structure of the data domain.

Time and vertex domains. Before introducing the time-vertex operators, we momentarily diverge by presenting the standard definitions in the time and graph domains. The main discrete calculus operator in time is the first order difference operator \( \mathbf{X} \nabla_T^-|_t = x_t - x_{t-1} \), taken here with periodic boundary conditions. Hence, the symmetric time Laplacian matrix \( \mathbf{L}_T = \nabla_T^2 \nabla_T \) is the discrete second order derivative in time with reversed sign \( \mathbf{X} \mathbf{L}_T|_t = -x_{t+1} + 2x_t - x_{t-1} \) and with \( x_{t+1} = x_1 \). As a circulant matrix, it has eigendecomposition \( \mathbf{L}_T = \mathbf{U}_T \mathbf{A}_T \mathbf{U}_T^T \), where \( \mathbf{A}_T(k,k) = 2(1-\cos(\omega_k)) \).

The operator corresponding to the time derivative in the vertex is the edge derivative. Given a graph signal \( \mathbf{x} \in \mathbb{R}^N \), the edge derivative with respect to edge \( e = (n, m) \) at vertex \( n \) is given by
\[
\frac{\partial \mathbf{x}}{\partial e}_n = \sqrt{\mathbf{W}(n,m)} [\mathbf{x}_n - \mathbf{x}_m].
\]
Therefore the graph gradient of \( \mathbf{x} \) at vertex \( n \) is
\[
\nabla_G \mathbf{x}_n = \left\{ \frac{\partial \mathbf{x}}{\partial e}_n \right\}_{e \in \mathcal{E}}
\]
and, as before, \( \mathbf{L}_G = \nabla_G^2 \nabla_G \), where \( \nabla_G^2 \) is the divergence operator of the graph.

Joint domain. We define the joint gradient of a time-vertex signal \( \mathbf{X} \) by concatenation of the joint time and graph gradients:
\[
\nabla_J \mathbf{x} = \text{vec} \left( \begin{bmatrix} \nabla_T \mathbf{X} \\ \nabla_G \mathbf{X} \end{bmatrix} \right).
\]
Therefore \( \nabla_J \) can be rewritten as
\[
\nabla_J = \begin{bmatrix} \nabla_T \otimes \mathbf{I}_G \\ \mathbf{I}_T \otimes \nabla_G \end{bmatrix}.
\]
The Laplacian is classically defined to equal the divergence of the gradient, and also in our case the joint Laplacian is \( \mathbf{L}_J = \nabla_J^2 \nabla_J \). Expanding the expression while exploiting the mixed-product property of the Kronecker product, we find
\[
\mathbf{L}_J = (\nabla_T \otimes \mathbf{I}_G)^* (\nabla_T \otimes \mathbf{I}_G) + (\mathbf{I}_T \otimes \nabla_G)^* (\mathbf{I}_T \otimes \nabla_G)
\]
\[
= (\nabla_T^2 \nabla_T) \otimes \mathbf{I}_G + \mathbf{I}_T \otimes (\nabla_G^2 \nabla_G)
\]
\[
= \mathbf{L}_T \otimes \mathbf{I}_G + \mathbf{I}_T \otimes \mathbf{L}_G = \mathbf{L}_T \times \mathbf{L}_G,
\]
and therefore \( \mathbf{L}_J \) is also equivalent to the Cartesian product between the time and graph Laplacians\(^1\). The result of the Cartesian product is a multilayer graph, referred to as the joint graph \( J \), where the original graph \( G \) is copied at each time step \( t = 1, 2, \ldots, T \). Additionally, each node at layer \( t \) is connected to itself at layer \( t-1 \) and \( t+1 \) with periodic boundary conditions. The Figure 1, which was borrowed from [16], shows an example of joint graph construction as Cartesian product.

\(^1\)In this work we consider the Cartesian product for its amenable spectral properties, but in general other graph products could be considered.
between a graph $G$ and the time domain represented as a cycle graph with length $T$.

It is useful to remind that the Kronecker product of the two eigenvectors basis $U_T$ and $U_G$ diagonalize the joint Laplacian with eigenvalues equal to the sum of all the pairs $(\omega_k, \lambda_k)$ [34]:

$$L_J = L_T \otimes I_G + I_T \otimes L_G$$

$$= (U_T A_T U_T^T) \otimes I_G + I_T \otimes (U_G A_G U_G^T)$$

$$= (U_T \otimes U_G)(A_T \times A_G)(U_T \otimes U_G)^* = U_J A_J U_J^*$$

where we have used the mixed-product property of the Kronecker product.

**Measures of joint variation.** The gradient and its various norms are often used as regularizers in regression because they capture the variation of the signal over a domain of interest. The $\ell_2$-norm of the joint gradient measures the total variation of the signal across edges and consecutive steps. Observe that

$$\| \nabla_J x \|_2^2 = x^T L_J x = \| \text{vec}(G_X) \|_F^2 + \| \text{vec}(T_X) \|_F^2$$

$$= \text{tr}(X^T L_G X) + \text{tr}(X L_T X^T)$$

meaning that $\| \nabla_J x \|_2^2$ is separable over the two domains.

Analogously, the $\ell_1$-norm of the joint gradient can be written as the sum of the $\ell_1$-norms

$$\| \nabla_J x \|_1 = \| \text{vec}(G_X) \|_1 + \| \text{vec}(T_X) \|_1,$$

which is often referred as the Total Variation (TV) norm. In general, it is possible to define a mixed norm $N_{p,q} (\cdot)$

$$N_{p,q} (x) \triangleq w_G \| \text{vec}(G_X) \|_p^p + w_T \| \text{vec}(T_X) \|_q^q$$

where the $p$-norm and the $q$-norm are computed independently on the two domains and $w_G, w_T$ are non-negative weights. Such norms are often useful when the signal vary differently (e.g., smooth or piece-wise) across the two domains, as we will show in Section VI-B.

### III. Dynamics over Graphs

This section motivates the JFT further by showing that it can be used to characterize two linear PDEs evolving over the graph by kernels defined in the joint frequency domain, and also to provide insight on standard non-linear PDEs used in epidemic modeling.

#### A. Linear dynamics on graphs

We are interested in linear PDEs whose solution at each time step can be expressed as a linear operator applied to the initial condition. In particular, we consider the heat diffusion and the wave equations defined in the discrete setting. We denote $x_1$ the initial condition of the PDEs, or equivalently in the joint spectral domain $Z(\ell, k) = \hat{x}_1(\ell) U_T^T(k, 1)$.

**Heat equation.** The discrete heat diffusion equation $x_{1} - x_{1-1} = -s L_G x_1$, is, arguably, one of the simplest dynamics described by differential equations. The parameter $s$ represents thermal diffusivity and is interpreted as a scale parameter for multiscale dynamic graph wavelet analysis [12] and graph scale-space theory [35]. It is well understood that $x_1 = (I - s L_G)^{-1} x_1$. Evaluating both the GFT and DFT, one also finds that the solution also has distinct structure in the joint spectral domain

$$\hat{X}(\ell, k) = \frac{1}{\sqrt{T}} a(\ell, \omega_k) - 1 Z(\ell, k)$$

where $a(\ell, \omega_k) = (1 - s \lambda_k) e^{-j \omega_k}$. The JFT of a heat diffusion process therefore exhibits a smooth non-separable low-pass form.

**Wave equation.** More interesting dynamics can be modeled by the discrete second order differential equation $X L_T = s L_G X$ represents a discrete wave propagating on a graph with speed $s > 0$. In the appendix we show that the solution in the joint spectral domain can be written as

$$\hat{X}(\ell, k) = \hat{K}(s, \omega_k) Z(\ell, k)$$

where

$$\hat{K}(s, \omega_k) = \sum_j \cos(t \theta_j) e^{-j \omega_k t}$$

where $\theta_j = \arccos(1 - s^2 \lambda_k)$ and to guarantee stability $s < 4/\lambda_{max}$. Therefore, in the joint spectral domain a wave has a simple, sparse, and distinctive form.

#### B. Complex dynamics over networks: the illustrative example of epidemic models

We next give an example of how the JFT provides insights on the evolution of a non-linear, discrete, and non-deterministic model for the spread of an infectious
disease. In particular, we focus on the dynamics corresponding to different compartmental models commonly used in epidemiology [36]. We simulated the epidemics spreading over \( N = 695 \) cities of Europe according to two different models: the Susceptible-Exposed-Infected-Recovered (SEIR) model and the SEIRS model, where the immunity of recovered individuals is only temporary. The models are parametrized by the contagion probability. The transform allows us to distinguish between the different parameters of the model.

Figure 2 shows the JFT of the simulated epidemic breakouts for the two different models. Each image shows the JFT of the signal describing the evolution of the number of infected over the graph. The transform is used to note that the joint spectral representation allows us to differentiate between the different parameters of the model.

**Fig. 2.** JFT of the signal representing the number of infected for different realization of epidemic spreads, using different models and contagion probability. The transform allows us to distinguish between the different parameters of the model.

**IV. FAST FILTERING OF TIME-VERTEX SIGNALS**

After recalling the definition of joint filters, we next present a novel algorithm to perform fast filtering on large graphs. Experiments illustrate that our algorithm achieves significantly better approximation of filtering objectives than state-of-the-art, while also not being constrained to a specific class of separable responses.

**A. Joint time-vertex filters**

A joint filter \( h(L_G, L_T) \) is a function defined in the joint two-dimensional spectral domain \( h : \mathbb{R}^+ \times \mathbb{R} \rightarrow \mathbb{C} \) and is evaluated at the graph eigenvalues \( \lambda \) and the angular frequencies \( \omega \). The output of a joint filter is

\[
h(L_G, L_T) x = U_J h(A_G, \Omega) U_J^T x,
\]

where \( h(A_G, \Omega) \) is a diagonal \( NT \times NT \) matrix defined as

\[
h(A_G, \Omega) = \text{diag} \left( \begin{array}{ccc}
    h(\lambda_1, \omega_1) & \cdots & h(\lambda_1, \omega_T) \\
    \vdots & \ddots & \vdots \\
    h(\lambda_N, \omega_1) & \cdots & h(\lambda_N, \omega_T)
\end{array} \right)
\]

and the \( \text{diag}(A) \) operator creates a matrix with diagonal elements the vectorized form of \( A \).

**Illustration: dynamic mesh filtering.** Figure 3 shows an example of joint filtering of a mesh representing a dancer\(^2\). The time-vertex signal \( X \in \mathbb{R}^{N \times T \times 3} \), with \( N = 1502 \) and \( T = 573 \), describes the time-varying coordinates of each vertex of the mesh. We design (a) a joint separable lowpass filter that attenuates high frequency components in both graph and time domains, and (b) a wave filter whose frequency response is described in Eq. (39). In the first case, we obtain the approximate skeleton of the mesh with rigid movements, whereas the wave filter produce a fluid (wavy) dancer, enhancing the frequency components in a non-linear fashion. We remark that this effect can only be obtained using non-separable filters.

\(^2\)http://research.microsoft.com/en-us/um/redmond/events/geometrycompression/data/default.html
Separable vs. non-separable filters. A notable family of joint filters are those that have separable response
\( h(\lambda, \omega) = h_1(\lambda)h_2(\omega) \). However, due to their simple form, separable filters cannot model the dynamics of PDEs (e.g., waves or heat diffusion), where there is an interplay between the temporal and graph frequency domains. For this reason, in the following we aim to find an efficient joint filter implementation for separable as well as non-separable filtering objectives.

B. Fast joint filtering

Due to the high complexity of eigendecomposition, graph filters are almost always implemented using fast 2D polynomial [37] and rational [28] approximations. In the context of time-vertex analysis, the importance of fast joint filtering is emphasized by the increase of the problem’s dimensions. Recognizing this need, researchers have recently proposed distributed joint filter 2D Chebyshev polynomial [16] and separable rational [17] implementations, appropriate for arbitrary and separable joint response functions, respectively. In the following, we improve upon state of the art by enhancing the filtering approximation at a similar (up to logarithmic factors) complexity.

The Fast Fourier Chebyshev (FFC) algorithm. The basic idea of our algorithm is to exploit the small complexity of FFT and perform graph filtering in the time-frequency domain. Concretely, to filter \( X \) with response \( h(\lambda, \omega) \), we do the following:

1. Compute the FFT of every row of \( X \), at a total complexity of \( O(NT \log T) \).
2. For each \( \omega_k \), approximate \( h(\lambda, \omega_k) \) with a Chebyshev polynomial of order \( M_G \) and use the fast graph Chebyshev recursion [11] to filter the corresponding angular frequency component of \( X \). The complexity of this step is \( O(M_GT|\mathcal{E}|) \).
3. Use the inverse FFT to obtain the filtered time-vertex signal, with complexity \( O(NT \log T) \).

Our scheme can approximate both separable or non-separable joint filters using \( O(T|\mathcal{E}|M_G + NT \log T) \) operations, which up to a logarithmic factor is a linear complexity to the number of edges \( |\mathcal{E}| \), nodes \( N \), timesteps \( T \), and filter order \( M_G \). Moreover, it can be performed distributedly since both the FFT and the graph Chebyshev recursion necessitate only local or few hop information.

Numerical comparison. To evaluate the approximation properties of the above scheme, we show in Figure 4 numerical experiments for an ideal separable lowpass filter and a non-separable wave filter on a time-vertex graph with size \( N = 5000, T = 3000 \). In detail, the approximated filtering functions (low pass and wave) are, respectively,

\[
\begin{align*}
    h_{\text{LP}}(\lambda_t, \omega_k) &= e^{-|\lambda_t - \lambda_{cf}|} - e^{-(|\omega_k| - \omega_{cf})} \\
    h_{\text{wave}}(\lambda_t, \omega_k) &= e^{-|\pi|\omega_k| - \arccos(1-\lambda_{cf}^2M_{max})^2}
\end{align*}
\]

For each case, we compare our algorithm with the state-of-the-art, i.e., Chebyshev2D approximation [16] of complexity and the ARMA2D approach [17], while choosing \( M_G \) and \( M_T \) as graph and temporal polynomial orders, respectively (here \( M_G = M_T \)).

As shown in Figure 4, FFC results in a significant improvement in accuracy for the same order and the difference is particularly prominent in the non-separable case (ARMA2D cannot be used here). We remark however that, to interpret these results correctly, one has to consider the complexity of each method:

<table>
<thead>
<tr>
<th>method</th>
<th>complexity</th>
<th>applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td>FFC</td>
<td>( O(T</td>
<td>\mathcal{E}</td>
</tr>
<tr>
<td>Cheby2D [16]</td>
<td>( O(T</td>
<td>\mathcal{E}</td>
</tr>
<tr>
<td>ARMA2D [17]</td>
<td>( O(T</td>
<td>\mathcal{E}</td>
</tr>
</tbody>
</table>

Therefore, for the same order, the three different methods feature slightly different complexities, implying that a direct comparison of accuracy is very difficult. Figure 4, which compares error as a function of the order, does not reveal that the three different methods actually have different computation times for the same order. Nevertheless, in our experiments for all orders larger than 2, the asymptotic complexity of FFC is the smallest (since here \( M_G = M_T \), \( \log T < M_T M_G \), and \( \log T < |\mathcal{E}|M_T / N \)). We also note that, in practice one often needs \( M_T \gg \log T \) to achieve a good approximation, in which case FFC is the fastest.
V. TIME-VERTEX DICTIONARIES AND FRAMES

So far, we have looked at time-vertex signals through the lenses of the canonical joint Fourier bases. However, in some cases it is beneficial to also consider alternative representations. For example, in the classic case, the wavelet and the short time Fourier transforms respectively enable time-scale and time-frequency analysis of the signal. The purpose of this section is to show how one can define analogous representations for time-vertex signals. These can be used for instance to generate features given as an input to a classifier (see Section VI-C) or to regularize an optimization problem such as (34) in Section VI-C.

Classically, the atoms of the representations are built by applying a transform (e.g. scaling or modulation) to a mother function and shifting the resulting functions. We follow a similar approach, with the difference that the mother function is replaced by a kernel defined in the time-vertex frequency domain and the shifting has to be replaced by an operator suitable to graphs. The spectral time-vertex wavelet and the short time vertex Fourier transforms follow as consequences of our framework.

A. Joint time-vertex localization

The ability to localize a kernel over a particular time and vertex is a key ingredient of our dictionary construction. In the following, we derive such a joint localization operator as a generalization of the graph localization operator [9], [33], [38], which localizes a kernel \( h(L_G) \) onto vertex \( v_m \)

\[
\mathcal{T}^G_m h = h(L_G) \delta_m = \sum_{\ell=1}^{N} h(\lambda_{\ell}) u_{\ell}(m) u_{\ell}.
\] (19)

Above, \( \delta_m \) is a Kronecker delta centered at vertex \( v_m \). Similarly, in the joint domain we define the joint time-vertex localization operator as the filtering with a two-dimensional Kronecker delta

\[
\mathcal{T}^J_{m,\tau} h \triangleq h(L_G, L_T) (\delta_m \otimes \delta_{\tau}).
\] (20)

It turns out that the joint time-vertex localization operator has the advantages of both the graph localization and the traditional translation operator. Indeed, we observe the following relations

\[
\mathcal{T}^J_{m,\tau} h(n,t) = \frac{1}{T} \sum_{\lambda=1}^{N} h(\lambda, \omega_k) \bar{u}_{\ell}(m)e^{-j\omega_k \tau} u_{\ell}(n)e^{j\omega_k t}
\]

\[
= \frac{1}{T} \sum_{k=1}^{T} \left( \sum_{\ell=1}^{N} h(\lambda_{\ell}, \omega_k) \bar{u}_{\ell}(m) u_{\ell}(n) \right) e^{j\omega_k (t-\tau)}
\] (21)

From (21), it follows that joint localization consists of three steps: (a) localizing independently all kernels \( h(\cdot, \omega_k) \), (b) computing the inverse DFT along the other dimension, and (c) translating the result. Joint localization is thus equivalent to independent application of a graph localization and a translation.

B. Joint time-vertex dictionaries

We proceed to present our dictionary construction for time-vertex signals. We start with a mother time-vertex kernel \( h(\lambda, \omega) \) and a transformation \( s_{z_\lambda, z_\omega} (\cdot, \cdot) \) parametrized by some values \([z_\lambda, z_\omega] \) belonging to the finite 2D set \( Z_\lambda \times Z_\omega \subset \mathbb{R}^2 \) and controlling the kernel’s shape along the vertex and time domains. The transformed kernel is then obtained by composition

\[
h_{z_\lambda, z_\omega}(\lambda, \omega) = h(s_{z_\lambda, z_\omega}(\lambda, \omega)).
\] (22)

We build our dictionary by transforming \( h(\lambda, \omega) \) with all \( z_\lambda, z_\omega \in Z_\lambda \times Z_\omega \), (possibly) normalizing, and jointly localizing the resulting kernels \( h_{z_\lambda, z_\omega}(\lambda, \omega) \) at each node \( m \) and time \( \tau \). Concretely, the dictionary is

\[
\mathcal{D}_h = \{ \mathcal{T}^J_{m,\tau} h_{z_\lambda, z_\omega} \} \quad \text{for} \quad m \in \mathcal{V}, \quad \tau = 1,2,\ldots,T,
\]

\[
\quad \text{and} \quad [z_\lambda, z_\omega] \in \{ Z_\lambda \times Z_\omega \}.
\] (23)

When \( \mathcal{D}_h \) is overly redundant, one may choose to consider only a subset of values for \( m \) and \( \tau \).

We next consider two interesting examples of the proposed dictionary construction that are generalizations of the short-time Fourier and wavelet transforms [39], [40]:

Short Time-Vertex Fourier Transform (STVFT). Set \( s_{z_\lambda, z_\omega} (\cdot, \cdot) \) to a shift in the spectral domain

\[
s_{z_\lambda, z_\omega}(\lambda, \omega) = [\lambda - z_\lambda, \omega - z_\omega].
\] (24)

This transform can be considered as a modulation. Nevertheless, we note that it does not correspond to a multiplication by an eigenvector as in [33], [41]. Our construction is more related to [42, Section 3]. Then, given a separable mother kernel \( h(\lambda, \omega) = h(\lambda) h_T(\omega) \) and a finite 2D set \( Z_\lambda \times Z_\omega \subset \mathbb{R}^2 \), the STVFT of signal \( X \) is defined as

\[
\text{STVFT}\{X\}(m,\tau, z_\lambda, z_\omega) \triangleq (X, \mathcal{T}^J_{m,\tau} h(\lambda - z_\lambda, \omega - z_\omega))
\]

\[
= \frac{1}{\sqrt{T}} \sum_{\ell,k} h(\lambda_{\ell} - z_\lambda, \omega_{k} - z_\omega) \bar{\mathcal{X}}(\ell, k) u_{\ell}(m) e^{j\omega_{k} \tau}.
\]

Provided that \( h(\lambda, \omega) \) is localized around \([0,0]\), the amplitude of the coefficient \((m, t, z_\lambda, z_\omega)\) indicates the presence of the spectral mode \([z_\lambda, z_\omega] \) at vertex \( m \) and time \( \tau \). Moreover, since the mother kernel is separable, the design in the two domains can be performed independently:

For the graph domain, we suggest to select the values of \( z_\lambda \) to be equally spaced in \([0, \lambda_{\max}] \) [42,
The spacing should be selected such that $\sum_{z_\lambda \in \mathbb{Z}_\omega} h_T^\dagger(\lambda - z_\lambda) \approx c$ for every $\lambda$, ensuring good conditioning of the associated frame (see [11, Theorem 5.6]). Because of the graph irregularity, in most of the cases, we need to keep all possible values for $m$, i.e., $m = 1, 2, \ldots, N$.

For the time domain, we recover a traditional STFT, with the difference that $h_T(\omega)$ is defined in the spectral domain. Nevertheless, for convenience, the window can still be designed in the time domain. As a rule of thumb $|\mathbb{Z}_\omega| = l_{h_T}$, where $l_{h_T}$ is the support of $h_T$ in the time domain\(^3\), and the values of $\tau$ should be sampled regularly with a spacing $\frac{l_{h_T}}{N}$, where $R$ is the desired redundancy. The choice of the window is highly linked to the addressed problem. Nevertheless, there are some general rules that are usually considered. Because of the Heisenberg uncertainty principle, there is a trade-off between localization in time and frequency. As a result, one should choose first the desired precision in one of the two domains. Then, a window shape can be selected that minimizes the uncertainty in order to maximize the simultaneous precision in both domains (i.e., it should be similar to a Gaussian function). Eventually, the sampling parameters have to be chosen. Usually in order to avoid unnecessary computations, the length of the support of the windows is used as the number of frequency channels. Then, the shift in time is selected to achieve a desired redundancy or to ensure some properties such as tightness. For a more complete treatment we refer the reader to [39].

Spectral Time-Vertex Wavelet Transform (STVWT). Following the idea developed in [11], we set $s_{z_\lambda, z_\omega}(\cdot, \cdot)$ to a generalized graph dilation (or scaling), i.e., a multiplication in the spectral domain

\[
s_{z_\lambda, z_\omega}(\lambda, \omega) = [z_\lambda \lambda, z_\omega \omega].
\]

Then, given a kernel $h(\lambda, \omega)$ the STVWT of $X$ reads

\[
\text{STVWT}\{X\}(m, \tau, z_\lambda, z_\omega) \triangleq \langle X, T_{m, \tau}^J h(z_\lambda \lambda, z_\omega \omega) \rangle = \frac{1}{\sqrt{T}} \sum_{\ell, k} h(z_\lambda \lambda, z_\omega \omega) \hat{X}(\ell, k) u_\ell(m) e^{j\omega_k \tau},
\]

where $z_\lambda, z_\omega$ are the scale parameters for the vertex and the time dimensions. A usual requirement for $h(\lambda, \omega)$ is that it has a zero DC component, i.e., $h(0, 0) = 0$. Contrarily to the STVFT, the mother kernel here may not be separable, as illustrated in VI-C. The choice of the discretization lattice $[m, \tau, z_\lambda, z_\omega]$ is thus more involved and case dependent: we suggest that $m$ and $\tau$ take all possible $N$ and $T$ values respectively, while $z_\lambda$ and $z_\omega$ are carefully selected depending on the application.

This choice is justified by the computational complexity detailed in the following.

C. Joint time-vertex frames

To make the proposed dictionaries and associated signal representations usable in practice, we next provide answers to three key questions: (a) How can we compute the representations efficiently (i.e., performing analysis and synthesis)? (b) How can we guarantee that the associated transforms are well conditioned such that they can be successfully inverted? (c) How to efficiently invert them, recovering the original signal? The second point is particularly important since a well conditioned transform allows for more robust representations, for instance when the dictionary is used to solve a synthesis or analysis regression problem with a sparse regularizer.

Efficient analysis and synthesis. The dictionary atoms can be seen as a filter-bank $\{h_z(\lambda, \omega)\}_{z \in \mathbb{Z}_\lambda \times \mathbb{Z}_\omega}$, in which case the operators going from the signal to the representation domain and back are the analysis operator

\[
D^h\{X\}(m, \tau, z) = \langle X, T_m^J h_z \rangle = C_z(m, \tau),
\]

and the synthesis operator

\[
D^h\{C\}(n, t) = \sum_z (C_z, T_n^J h_z) = Y(n, t).
\]

Notice that in general $X \neq Y$, and equality holds only when the filter-bank is a unitary tight frame.

Instead of computing the dictionary explicitly (an operation that is costly both in memory and in computations), one may acquire the analysis coefficients for all $m, \tau$ by joint filtering $X$ with kernel $h_z$ taking advantage of the relation $C_z = \text{mat}(h_z(L_G; L_T) x)$. Similarly synthesis can be performed by summing filtering operations. Using our FPC filtering algorithm presented in Section IV-B, the total analysis complexity is thus $O(|\mathbb{Z}_\lambda \times \mathbb{Z}_\omega|(T|E| M_G + NT \log T))$, where typically $M_G \approx 50$.

Conditioning and frame bounds. In several applications in signal processing one is interested not only in processing data in another convenient representation, but also to recover the original signal from its alternative representation. Redundant invertible dictionaries are referred to as frames [39], [43]. The following theorem generalizes the classic [39] results regarding the frame bounds, providing a condition for a joint time-vertex dictionary to be a frame, as in the case of graphs [42, Lemma 1], [11, Theorem 5.6].

**Theorem 1.** Let $\{h_z(\lambda, \omega)\}_{z \in \mathbb{Z}_\lambda \times \mathbb{Z}_\omega}$ be the kernels of a time-vertex dictionary $D^h$, and set

\[
A = \min_{l, k} \sum_z |h_z(\lambda_l, \omega_k)|^2, \quad B = \max_{l, k} \sum_z |h_z(\lambda_l, \omega_k)|^2.
\]

\[\sum_{z} |h_z(\lambda_l, \omega_k)|^2 \approx c \quad \text{for every } \lambda_l, \omega_k, \text{ ensuring good conditioning of the associated frame,}
\]

\[\sum_{z} |h_z(\lambda_l, \omega_k)|^2 \approx c \quad \text{for every } \lambda_l, \omega_k, \text{ ensuring good conditioning of the associated frame,}
\]

\[\sum_{z} |h_z(\lambda_l, \omega_k)|^2 \approx c \quad \text{for every } \lambda_l, \omega_k, \text{ ensuring good conditioning of the associated frame,}
\]
If $0 < A \leq B < \infty$, then $D_h$ is a frame in the sense:

$$A\|X\|_F^2 \leq \|D_h\{X\}\|_F^2 \leq B\|X\|_F^2$$

(26)

for any time-vertex signal $X \in \mathbb{R}^{N \times T}$.

The proof can be found in an earlier version of this work [18, Theorem 1].

The theorem asserts that, if $A > 0$, no information is lost when the analysis operator is applied to a time-vertex signals, thus the transform is invertible. Furthermore, the ratio $A/B$ of the frame bounds is related to the condition number of the frame operator $S_h\{X\} = D_h^\dagger\{D_h\{X\}\}$, hence it is decisive for efficient reconstruction when we want to recover the signal from its representation solving an optimization problem [11, Section 7].

**Efficient inversion.** To recover the signal $X$ from the coefficients $C$, a solution is to use the pseudo-inverse, i.e. $X = D_h^\dagger\{C\}$ or to solve the following convex problem $\arg\min_{X} \|D_h\{X\} - C\|_F^2$. Problematically, these are computationally intractable for large value of $N$ and $T$. We will instead design a dual set of kernels that allows us to invert the transform by a single synthesis operation. To this end, we search for a set of filters $\tilde{h}$ such that $D_h^\dagger\{D_h\{X\}\} = X$. It is not difficult to see that this equality is satisfied when

$$\sum_{z_3, z_ω} \tilde{h}_{z_3, z_ω}(\lambda_ℓ, ω_k) h_{z_3, z_ω}(\lambda_ℓ, ω_k) = 1, \quad \forall \lambda_ℓ, ω_k.$$

(27)

Although redundant joint time-vertex frames admit an infinite number of dual kernel sets satisfying (27), the typical choice is to use the canonical dual, defined as

$$\tilde{h}_{z_3, z_ω}(\lambda_ℓ, ω_k) = \left( \sum_{z_3', z_ω'} h_{z_3', z_ω'}^2(\lambda_ℓ, ω_k) \right)^{-1} h_{z_3, z_ω}(\lambda_ℓ, ω_k).$$

(28)

In fact, this corresponds to the pseudo-inverse of $D_h$, i.e., $D_h^\dagger = D_h^\ast$, while also having a low computational complexity.

To summarize, given an invertible time-vertex transform $D_h$ and coefficients $C$, the inverse transform of $D_h$ associated with the set of kernels $\{\tilde{h}_{z_3, z_ω}\}_{z_3, z_ω} \in \mathcal{Z}$ is

$$X = D_h^\dagger\{C\} = \sum_{z_3, z_ω} \tilde{h}_{z_3, z_ω}(L_G, L_T) C_{z_3, z_ω},$$

(29)

where $\tilde{h}$ is defined in (28).

**VI. EXPERIMENTS**

The suitability of the time-vertex framework for several classes of problems is illustrated on a wide variety of datasets: (a) dynamic meshes representing a walking dog and a dancing man, (b) the Caltrans Performance Measurement System (PeMS) traffic dataset depicting high resolution daily vehicle flow of 10 consecutive days in the highways of Sacramento measured every 5 minutes, (c) simulated SEIR- or SEIRS-type epidemics over Europe, (d) the Kuala Lumpur City Centre (KLCC) time-lapse video and (e) earthquake waveforms recorded by seismic stations geographically distributed in New Zealand, connected to the GeoNet Network.

Results suggest that joint analysis of time-vertex signals can bring forth benefits in signal denoising and recovery, learning and source localization problems. We remark that all the experiments were done using the GSPBOX [44], the UNLocBoX [45] and the LT-FAT [46]. Code reproducing the experiments is available at https://lts2.epfl.ch/reproducible-research/a-time-vertex-signal-processing-framework/.

A. **Compactness of representation**

A key motivation behind the joint harmonic analysis is the capability of encoding time-varying graph-dependent signal evolution in a compact way. Our first step will therefore be to examine the energy compaction of the JFT transform in four datasets: two meshes representing a dancer ($N = 1502$ points in $\mathbb{R}^3$ and $T = 570$ timesteps) and a dog walking ($N = 2502$ points in $\mathbb{R}^3$ and $T = 52$ timesteps), the PeMS traffic flow dataset ($N = 710$ stations measuring traffic over $T = 2880$ intervals of 5-minute length each) and the number of infected individuals in an SEIRS epidemic (see Section III-B for description). In both dynamic meshes, we consider three signals corresponding to each
of the \((x, y, z)\) coordinates associated with each point. Transforms with good energy compaction are desirable because they summarize the data well and can be used to construct efficient regularizers for regression problems.

To measure energy compactness, we compute the DFT, GFT and JFT for each dataset, we replace the spectrum coefficients with magnitudes smaller than the \(p\)-th percentile with zeros and perform the corresponding inverse transform on the resulting coefficients. Denoting by \(\mathbf{X}\) the original signal and \(\mathbf{X}_p\) the compressed one, the compression error is for each \(p\) given by \(\|\mathbf{X}_p - \mathbf{X}\|_F / \|\mathbf{X}\|_F\). As shown in Figure 5, JFT exhibits better energy compaction properties in all the datasets, and especially for the meshes where the graph captures well the signal structure.

### B. Regression problems with joint variation priors

We next examine the utility of joint variation priors for regression problems in two example applications.

**Denoising of dynamic meshes.** Whenever a smoothness prior can be assumed, the joint Tikhonov regularization can be used to denoise a time-varying graph signal. The prior can be easily expressed in the time-vertex domain thanks to Eq. (11). Joint denoising is then performed by solving the following optimization problem

\[
\arg\min_{\mathbf{X}} \|\mathbf{X} - \mathbf{Y}\|^2_F + \tau_1 \|\nabla_G \mathbf{X}\|^2_F + \tau_2 \|\mathbf{X} \nabla_T\|^2_F, \tag{30}
\]

where the regularization terms require the solution to be smooth in both graph and time domains. This problem has a closed form solution in our framework, which is a joint non-separable lowpass filter

\[
h_{\text{TIK}}(\lambda_\ell, \omega_k) = \frac{1}{1 + \tau_1 \lambda_\ell + 2 \tau_2 (1 - \cos(2\pi \omega_k))}. \tag{31}
\]

In order to investigate its performance, we consider the vertices of a 3D time-varying mesh of size \(2502 \times 59 \times 3\) (vertex \(\times\) time \(\times\) dimension) representing a walking dog, we add Gaussian noise to the coordinates, we build a \(k\) nearest neighbor graph based on the distances between the time average of the coordinates and finally we solve the problem (30) for each coordinate dimension. We averaged the results over 20 realizations of the noise.

The meshes in Figure 6 represent, from left to right, the original, the noisy and the recovered one, for one realization of the noise. Remarkably, the normalized error drops from 0.20 to 0.06, respectively before and after denoising, making the dog distinguishable again. As side effect, the dog appears to be thinner, due to the graph regularization. The heat-map in the left corner of the figure shows the role of the regularization parameters.

We found (using exhaustive search) that the lowest error is achieved when \(\tau_1 = 0.71\) and \(\tau_2 = 1.78\). We compare the performance of the joint Tikhonov regularization with respect to time- and graph-only for the best parameter combinations of all methods. The boxplot on the right shows the minimum achievable error statistics in the three cases over the 20 realizations. It is easy to see that the graph plays a major role in the denoising, since it encodes the structural information of the mesh. Nevertheless, the joint approach performs the best, i.e., 0.062 ± 0.0002, taking advantage of the smoothness in both domains, while graph and time methods achieve 0.067 ± 0.0003 and 0.095 ± 0.0002, respectively.

**Inpainting of time-lapse video.** We consider the problem of time-vertex signal recovery from noisy, corrupted, and incomplete measurements. Depending on the characteristics of the signal, the prior \(N_{p,q}(x)\) with different values of \(p\) and \(q\) and different weights can be used. A typical signal recovery problem in signal processing is the image inpainting, i.e., trying to replace corrupted or lost part of the image. Since patch-graphs allow non-local image processing [47], our goal is to extend graph-based non-local processing to video inpainting and recovery. However, since our framework is constrained to static graphs, we focus on the particular case of time-lapse videos, whose structure stays majorly invariant throughout the video.

To this end, we corrupted a time-lapse video that shows the skyline of the Kuala Lumpur City Centre, which statistical properties were amenable from a graph perspective, being the skyline static with time-varying colors. The video has size \(160 \times 214 \times 3 \times 604\) (height \(\times\) width \(\times\) colors \(\times\) frames). We removed 20% of the pixels and 20% of the frames from the original KLCC video, achieving a normalized error of 0.61. The
Inpainting is performed solving the optimization problem for each color using as regularizer $N_{1,2}(\mathbf{x})$:

$$\arg \min_{\mathbf{x}} \| \mathbf{M} \odot \mathbf{X} - \mathbf{Y} \|_F^2 + \gamma_1 \| \nabla_G \mathbf{X} \|_1 + \gamma_2 \| \mathbf{X} \nabla_T \|_F^2,$$

(32)

where $\mathbf{M}$ is the mask of the missing entries. The patch graph $G$ is constructed from the video averaged in time. The rationale is that the $l_1$ over the graph will restore the missing pixels, being each frame approximately piecewise constant, whereas the $l_2$ norm in time recovers the smooth changes of the colors from dawn until dusk.

**TABLE I**

<table>
<thead>
<tr>
<th>Regularizer</th>
<th>Pixels</th>
<th>Frames</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tikhonov</td>
<td>0.051</td>
<td>0.100</td>
<td>0.059</td>
</tr>
<tr>
<td>TV</td>
<td><strong>0.048</strong></td>
<td>0.122</td>
<td>0.060</td>
</tr>
<tr>
<td>$N_{1,1}(\mathbf{x})$</td>
<td>0.056</td>
<td>0.059</td>
<td>0.057</td>
</tr>
<tr>
<td>$N_{1,2}(\mathbf{x})$</td>
<td>0.050</td>
<td><strong>0.055</strong></td>
<td><strong>0.051</strong></td>
</tr>
<tr>
<td>$N_{2,1}(\mathbf{x})$</td>
<td>0.061</td>
<td>0.103</td>
<td>0.068</td>
</tr>
<tr>
<td>$N_{2,2}(\mathbf{x})$</td>
<td>0.053</td>
<td>0.066</td>
<td>0.055</td>
</tr>
</tbody>
</table>

We compare the recovery performance with all the joint regularizers $N_{p,q}(\mathbf{x})$ for $p,q \in \{1,2\}$, and with two baseline algorithms, based on 3D-Tikhonov and isotropic 3D-TV regularizations [48]. The last two correspond to using a grid graph with equal weights on the edges. Table I reports the normalized errors averaged over the pixels-only, frames-only and the whole video. The better performance achieved by the joint regularizer $N_{1,2}(\mathbf{x})$ is due to its capability to restore missing frames, while missing pixels recovery performances are almost the same. Figure 7 illustrates a summary statistics of the errors computed over each frame. Although TV performs the best in the median, in case of occluded frames the error is much larger w.r.t. the joint recovery, leading to a higher average error.

**C. Overcomplete representations**

Last, we examine the utility of STVFT and STVWT, respectively, as a feature extractor for dynamic mesh clustering and as a dictionary used to uncover the wave-like structure and epicenter of a seismic event.

**Clustering dynamic meshes using STVFT.** We consider the motion classification of a dynamic mesh representing the dancer, corrupted with additive sparse noise with density 0.1 with normally distributed entries and SNR of $-20$ dB and $-10$ dB. Our objective is to determine the phase of the dance (moving arms, stretching legs and bending body) at each frames by performing spectral clustering on some representation of the windowed signal. To obtain the ground-truth, we labeled each frame by hand and verified that, when the noiseless signal (i.e., the actual trajectory of the points in time over each window) was used to define the features, one obtains a classification accuracy of 0.926.

Since we want to localize spatial-structured phenomena in time, our approach will be to use a STVFT to derive the representation. To capture the geometry of the problem, we used a nearest neighbour graph constructed based on the coordinates of the mesh vertices averaged in time; this graph was fixed for the whole sequence. As explained in Section V-B, the STVFT is separable, meaning that we can handle the vertex and the time dimensions separately. In the time domain we use a rectangular window with support equal to 50 samples in time and spacing such that the overlap is 60%. For the vertex dimension, we use the an Iersine kernel (defined in the GSPBOX [44]) that we uniformly translate at 5 different positions in the graph spectral domain.

The STVFT provides features associated to a time instant that we can directly use to classify the dance (see Figure 8 for a visual illustration of the clustering results). Other transforms such as GFT, DFT, and JFT

![Image](https://via.placeholder.com/150)
do no have this property. Hence, in order to compare with these other transforms, we use the same rectangular windows (width 50, overlap 20 samples) to extract 27 time sequences from the signal. We then used the transformed data associated with each sequence as a point to be clustered.

Figure 9 illustrates the clustering accuracy statistics over 20 realizations of sparse noise for features constructed based on the magnitude of five representations: the windowed sequences, as well as their GFT, DFT, JFT, and STVFT representations. Observe how the presence of sparse noise severely hampers classification when the raw signal is used, with the average accuracy dropping from 0.926 for the clean signals to 0.469 and 0.74 for $-20 \text{ dB}$ and $-10 \text{ dB}$, respectively.

We can also see that the two representations leading to the highest median accuracy are the JFT and STVFT, suggesting the utility of joint harmonic representations. Nevertheless, the STVFT provides more robust estimates with an average accuracy of 0.869 rather than 0.792 for the JFT at $-20 \text{ dB}$.

**Seismic epicenter estimation with STVWT.** We analyze seismic events recorded by the GeoNet sensor network whose epicenters were chosen to be randomly distributed in different areas of New Zealand. We extend the results presented in [18] to a greater dataset using the STVWT with mother kernel based on the wave PDE, which allows us to decompose the signal as sum of PDE solutions. As a first approximation, when the waves propagate in a continuous domain or a regular lattice, seismic waveforms can be modeled as oscillating damped waves [49]. Our premise is thus that we can approximate the seismic waveforms using a small set of damped waves propagating on the graph connecting the seismic stations. Thus, we expect the damped wave mother kernel

$$h(\lambda_t, \omega_k) = \frac{1}{\sqrt{T}} \frac{e^{\beta + j\omega_k + \lambda_t/2 - 1}}{2\cosh (\beta + j\omega_k) + \lambda_t/2 - 1} \tag{33}$$

to be a good approximation of the seismic waves recorded by the sensors, with the damping factor $\beta$ chosen to fit the damping present in the seismic signals. To construct the STVWT, we select 10 equally spaced values in $[0, 2]$ for $z_\lambda$ and set $z_\omega = 1$. To estimate the epicenter of the earthquake we solve

$$\arg \min_C \|D_h^* \{C\} - X\|_2^2 + \gamma \|C\|_1, \tag{34}$$

where $\gamma$ is the regularization parameter controlling the trade-off between the fidelity term (selected using exhaustive search) and the sparseness assumption and $D_h^*$ is the synthesis operator associated with STVWT.

The solution provides important pieces of information. Firstly, using the synthesis operator we can obtain a denoised version of the original process. Secondly, the non-zero coefficients of $C$, describe the origins and amplitudes of the different components. Therefore, for each node we take the maximum value over time and scales and we use the result as weights to average the graph coordinates to obtain the epicenter. Being a weighted average of different coordinates, the estimated source will not correspond to a vertex of the graph in general. We compare the performance of STVWT with the estimate obtained using only the signal amplitude: for each earthquake we average the coordinates of the stations using as weights energy of the signals. Figure 10 shows on the left the comparison over 40 different seismic events randomly distributed over the New Zealand between the two methods. STVWT based on the damped wave kernel achieves an average error of 48.5 km, providing an almost twofold improvement over the baseline, whose average performance is 88.3 km. On the right, it illustrates the estimate for 3 different seismic events and the respective seismic waveforms. These results show that the proposed method significantly improves the source estimation performance.

**VII. Conclusion**

This work puts forth a Time-Vertex Signal Processing Framework, that facilitates the analysis of graph structured data that also evolve in time. We motivate our framework leveraging the notion of partial differential equation on graphs. We introduce joint operators, such as time-vertex localization and we present a novel approach to significantly improve the accuracy of fast joint
filtering. We also illustrate how to build time-vertex dictionaries, providing conditions for efficient invertibility and examples of constructions. Our experimental results on a variety of datasets, suggest that the proposed tools can bring forth significant benefits in various signal processing and learning tasks involving time-series on graphs.

APPENDIX

A. Wave equation

In the continuous setting, the wave equation is

$$\partial_t^2 u - \Delta u = 0$$

where $u: \mathbb{R} \times \mathbb{R}^d \to \mathbb{C}$ is a function of both time $t \in \mathbb{R}$ and space $x \in \mathbb{R}^d$, with $\Delta$ being the Laplacian operator.

Assuming vanishing initial velocity $u(0) = 0$, the solution $u(t)$ is given via functional calculus by [50]

$$u(t) = \cos(t \sqrt{-\Delta}) u(0)$$

where $\cos(t \sqrt{-\Delta})$ is called propagator operator.

To obtain a discrete wave equation evolving on a graph, we approximate the second order time derivative with its stencil approximation and the continuous Laplacian $\Delta$ with the graph Laplacian $L_G$ with reversed sign:

$$X L_T = sL_G X,$$

where $s > 0$ is the speed of the propagation. We assume that the solution can be written as

$$x_t = K_s(L_G, t)x_1 = K_{t,s}x_1,$$

where $K_{t,s} = K_s(L_G, t)$ is a matrix obtained applying the function $K_s(L_G, t)$ to the scaled Laplacian $sL_G$ and parametrized by the time $t$. We will call the operator $K_{t,s}$ “the discrete analogue of the wave propagator” of Eq. (35). Therefore, matrix $X = [K_{t,s} x_1]^T = K_s x_1$ is obtained stacking the vectors $x_t$ of Eq. (37) along the columns. Substituting (37) into (36), we obtain

$$K_{t,s} L_T = sL_G K_{t,s} x_1$$

which in the graph spectral domain is

$$\tilde{K}_{t,s} \tilde{x}_1 = s \Lambda_G \tilde{K}_{t,s} \tilde{x}_1,$$

where $\tilde{K}_{t,s} = K_s(\Lambda_G, t)$. Equation (38) is formally analogous to the eigendecomposition of the operator $L_T$, therefore, the $\ell$-th row of $\tilde{K}_{t,s} \tilde{x}_1$ must be an eigenvector of $L_T$ with eigenvalue $\lambda_\ell$, for every $\ell$. Using Eq. (39), we obtain

$$K_s(\lambda_\ell, t) = \cos(t \theta_\ell),$$

with $\theta_\ell = \arccos(1 - \frac{\Lambda_G}{2})$. Since the $\arccos(x)$ is defined only for $x \in [-1, 1]$, to guarantee stability the parameter $s$ must satisfy $s < 4/\lambda_{\max}$. We remark that this result is in agreement with the stability analysis of numerical solver for the discrete wave equation presented in [50].

Taking the DFT of the wave kernel in Eq. (39), we obtain

$$\tilde{K}_s(\lambda_\ell, \omega_k) = \sum_t \cos(t \theta_\ell) e^{-j \omega_k t}$$

Therefore, the solution in the joint spectral domain can be written as

$$\tilde{X}(\ell, k) = \tilde{K}_s(\lambda_\ell, \omega_k) Z(\ell, k),$$

where $Z(\ell, k) = \tilde{x}_1(t) U^*_T(k, 1)$.

REFERENCES


V. Kalofolias, “How to learn a graph from smooth signals,” in *AISTATS*, 2016.

A. Loukas and D. Foucard, “Frequency Analysis of Temporal Graph Signals,” in *GLOBALSP*, 2016.


