

EIGENTRIADS AND EIGENPROGRESSIONS ON THE TONNETZ

Vincent Lostanlen

Music and Audio Research Lab, New York University, New York, NY, USA

ABSTRACT

We introduce a new multidimensional representation, named *eigenprogression transform*, that characterizes some essential patterns of Western tonal harmony while being equivariant to time shifts and pitch transpositions. This representation is deep, multiscale, and convolutional in the piano-roll domain, yet incurs no prior training, and is thus suited to both supervised and unsupervised MIR tasks. The eigenprogression transform combines ideas from the spiral scattering transform, spectral graph theory, and wavelet shrinkage denoising. We report state-of-the-art results on a task of supervised composer recognition (Haydn vs. Mozart) from polyphonic music pieces in MIDI format.

1. EIGENTRIADS

Let $\mathbf{x}[t, p] \in \mathcal{M}_{T,P}(\mathbb{R})$ the piano-roll matrix of a musical piece, either obtained by parsing symbolic data or by extracting a melody salience representation from audio [2]. The constant T (resp. P) is typically equal to 2^{10} (resp. 2^7). Within the framework of twelve-tone equal temperament, we define the major and minor triads as the tuples $\mathcal{I}_1 = (0, 4, 7)$ and $\mathcal{I}_0 = (0, 3, 7)$. For each quality $q \in \mathbb{Z}_2$ and frequency $\beta \in \mathbb{Z}_3$, let

$$\psi_{\beta_1, q}^{\text{triad}}[p] = \sum_{n=1}^3 \exp\left(2\pi i \frac{\beta n}{3}\right) \delta[p - \mathcal{I}_q[n]], \quad (1)$$

where $\delta[p - \mathcal{I}_q[n]]$ is the Kronecker delta symbol, equal to one if $p = \mathcal{I}_q[n]$ and zero otherwise. Let \mathcal{G}_q the induced subgraph of \mathcal{I}_q , where \mathcal{I}_q is understood as a set of vertices in \mathbb{Z}_P . Observe that $\{\psi_{\beta, q}^{\text{triad}}\}_{\beta}$ consists of the eigenfunctions of the unnormalized Laplacian matrix of \mathcal{G}_q :

$$\mathbf{L}_q^{\text{triad}}[p, p'] = |\mathcal{I}_q| \delta[p \in \mathcal{I}_q] \delta[p' \in \mathcal{I}_q] - \delta[p \overset{\mathcal{G}_q}{\sim} p'] \quad (2)$$

As a result, we propose to name *eigentriads* the complex-valued signals $\psi_{\beta_1, q}^{\text{triad}}$ in pitch space. We construct a multiresolution convolutional operator in the piano-roll domain by separable interference $\Psi_{(\alpha_1, \beta_1, q)}^{\text{triad}}[t, p] = \psi_{\alpha_1}[t] \psi_{\beta_1, q}^{\text{triad}}[p]$

This work is supported by the ERC InvariantClass grant. The author thanks Moreno Andreatta, Joanna Devaney, Peter van Kranenburg, Stéphane Mallat, Brian McFee, and Gissel Velarde for helpful comments.

between the aforementioned eigentriads and a family of temporal Gabor wavelets

$$\psi_{\alpha_1}[t] = \alpha_1 \exp\left(-\frac{\alpha_1^2 t^2}{2\sigma^2}\right) \exp(i\alpha_1 \xi t) \quad (3)$$

for $t \in \llbracket 0; T \rrbracket$. We set $\xi = \frac{2\pi}{3}$, $\sigma = 0.1$, $\xi = \frac{2\pi}{3}$, and $\log_2 \alpha_1 \in \llbracket 0; \log_2 T \rrbracket$. We define the *eigentriad transform* of \mathbf{x} as the rank-five tensor resulting from the complex modulus of all convolutions between \mathbf{x} and multivariable wavelets $\Psi_{(\alpha_1, \beta_1, q)}^{\text{triad}}[t, p]$:

$$\mathbf{U}_1(\mathbf{x})[t, p, q, \alpha_1, \beta_1] = \left| \mathbf{x} * \Psi_{(\alpha_1, \beta_1, q)}^{\text{triad}} \right| [t, p] \\ = \left| \sum_{t'=0}^{T-1} \sum_{p'=0}^{P-1} \mathbf{x}[t', p'] \Psi_{(\alpha_1, \beta_1, q)}^{\text{triad}}[t - t', p - p'] \right|, \quad (4)$$

where the difference in t (resp. in p) is computed in \mathbb{Z}_T (resp. in \mathbb{Z}_P). By averaging the tensor $\mathbf{U}_1(\mathbf{x})$ over the dimensions of time t , pitch p , and triad quality q , one obtains the matrix

$$\mathbf{S}_1(\mathbf{x})[\alpha_1, \beta_1] = \sum_{t \in \mathbb{Z}_T} \sum_{p \in \mathbb{Z}_P} \sum_{q \in \mathbb{Z}_2} \mathbf{U}_1(\mathbf{x})[t, p, q, \alpha_1, \beta_1]. \quad (5)$$

The operator $\mathbf{x} \mapsto \mathbf{S}_1(\mathbf{x})$ characterizes the relative amounts of ascending triads ($\beta_1 = 1$), descending triads ($\beta_1 = -1$), and perfect chords ($\beta_1 = 0$) at various temporal scales α_1 in the piece \mathbf{x} , while keeping a relatively low dimensionality, equal to $3 \log_2 T$. The averaging along variables t and p involved in Equation 5 guarantees that \mathbf{S}_1 is invariant to the action of any temporal shift operator $\tau_{\Delta t} : \mathbf{x}[t, p] \mapsto \mathbf{x}[t + \Delta t, p]$, as well as any pitch transposition operator $\pi_{\Delta p} : \mathbf{x}[t, p] \mapsto \mathbf{x}[t, p + \Delta p]$:

$$\forall \Delta t \in \mathbb{Z}_T, \forall \Delta p \in \mathbb{Z}_P, \mathbf{S}_1(\pi_{\Delta p} \circ \tau_{\Delta t} \circ \mathbf{x}) = \mathbf{S}_1(\mathbf{x}). \quad (6)$$

Furthermore, the averaging across triad qualities q implies approximate invariance to tonality, in the sense that replacing major triads by minor triads and vice versa in \mathbf{x} (insofar as this is feasible in the signal \mathbf{x} at hand) does not affect the matrix $\mathbf{S}_1(\mathbf{x})$. From the standpoint of serialist music theory [1], the presence of signed eigentriad frequencies $\beta_1 = \pm 1$ ensures that \mathbf{S}_1 is not invariant to retrogradation $\mathbf{R} : \mathbf{x}[t, p] \mapsto \mathbf{x}[-t, p]$, i.e. time reversal:

$$(\mathbf{R} \circ \mathbf{x}) \neq \mathbf{x} \implies \mathbf{S}_1(\mathbf{R} \circ \mathbf{x}) \neq \mathbf{S}_1(\mathbf{x}). \quad (7)$$

However, the averaging across triad qualities q causes \mathbf{S}_1 to be invariant to inversion $\mathbf{I} : \mathbf{x}[t, p] \mapsto \mathbf{x}[t, -p]$, i.e. reversal of the pitch axis:

$$(\mathbf{I} \circ \mathbf{x}) \neq \mathbf{x} \not\implies \mathbf{S}_1(\mathbf{I} \circ \mathbf{x}) \neq \mathbf{S}_1(\mathbf{x}). \quad (8)$$



The above property hinders the accurate modeling of chord progressions in the context of Western tonal music. Indeed, \mathbf{S}_1 fails to distinguish a perfect major cadence ($\mathbf{C}^{\text{maj}} \rightarrow \mathbf{F}^{\text{maj}}$) from a plagal minor cadence ($\mathbf{F}^{\text{min}} \rightarrow \mathbf{C}^{\text{min}}$), as one proceeds from the other by involution with \mathbf{I} . More generally, the eigentriad transform may suffice for extracting the quality of isolated chords (e.g. major vs. minor), but lacks longer-term context of harmonic tension and release so as to infer the tonal functions of such chords (e.g. tonic vs. dominant).

2. EIGENPROGRESSIONS

In this section, we introduce a second multidimensional feature \mathbf{U}_2 , built on top of \mathbf{U}_1 and named *eigenprogression transform*, that aims at integrating harmonic context in Western tonal music while still respecting the aforementioned requirements of invariance to global temporal shifts $\tau_{\Delta t}$ and pitch transpositions $\pi_{\Delta p}$. We begin by defining the Tonnetz as an undirected graph over the 24 vertices of triads $(p, q) \in \mathbb{Z}_{12} \times \mathbb{Z}_2$. Let $\mathcal{J}_1 = 4$ (resp. $\mathcal{J}_0 = 3$) the number of semitones in a major (resp. minor) third. The unnormalized Laplacian tensor of the Tonnetz is

$$\begin{aligned} \mathbf{L}^{\text{Tonnetz}}[p, q, q', p'] &= \left(\delta[(-1)^q(p - p') \in \mathcal{J}_q] \right. \\ &+ \delta[(-1)^{q'}(p - p') \in \mathcal{J}_{q'}] \\ &+ \delta[p - p'] \times \delta[q - q' + 1] \\ &\left. - 3\delta[p - p']\delta[q - q'] \right) \end{aligned} \quad (9)$$

We define eigenvalues λ of the Laplacian tensor, and corresponding eigenvectors \mathbf{v} , as the solutions of the following equation:

$$\begin{aligned} (\mathbf{L}^{\text{Tonnetz}} \overline{\otimes} \mathbf{v})[p, q] &= \sum_{p', q'} \mathbf{L}[p, q, q', p'] \mathbf{v}[p', q'] \\ &= \lambda \mathbf{v}[p, q]. \end{aligned} \quad (10)$$

The number of distinct eigenvalues $\lambda_1 \dots \lambda_K$ satisfying Equation 10 is equal to K , and their associated eigensubspaces form a direct sum: $\mathbf{V}_1 \oplus \dots \oplus \mathbf{V}_K = \mathcal{M}_{P,2}(\mathbb{R})$. For values of k such that $\dim \mathbf{V}_k = 1$, we define the k^{th} eigenprogressions as $\psi_k^{\text{Tonnetz}}[p, q] = \mathbf{v}_k[p, q]$, where $\mathbf{v}_k \in \mathbf{V}_k$ and $\|\mathbf{v}_k\|_2 = 1$. In contrast, for values of k such that $\dim \mathbf{V}_k = 2$, we arbitrarily select two vectors \mathbf{v}_k^{Re} and \mathbf{v}_k^{Im} satisfying $\mathbf{v}_k^{\text{Re}} \perp \mathbf{v}_k^{\text{Im}}$, $\|\mathbf{v}_k^{\text{Re}}\|_2 = \|\mathbf{v}_k^{\text{Im}}\|_2 = 1$, and $\text{span}(\{\mathbf{v}_k^{\text{Re}}, \mathbf{v}_k^{\text{Im}}\}) = \mathbf{V}_k$; and define eigenprogressions as

$$\psi_{\beta_2}^{\text{Tonnetz}}[p, q] = \mathbf{v}_k^{\text{Re}}[p, q] + i\mathbf{v}_k^{\text{Im}}[p, q]. \quad (11)$$

We define multivariable eigenprogression wavelets as

$$\Psi_{(\alpha_2, \beta_2, \gamma_2)}^{\text{prog}}[t, p, q] = \psi_{\alpha_2}[t] \psi_{\beta_2}^{\text{Tonnetz}}[p, q] \psi_{\gamma_2}^{\text{spiral}}[p], \quad (12)$$

where ψ_{α_2} is a temporal Gabor wavelet of frequency and $\psi_{\gamma_2}^{\text{spiral}}$ is a Gabor wavelet on the Shepard pitch spiral [4]:

$$\psi_{\gamma_2}^{\text{spiral}}[p] = \gamma_2 \exp\left(-\frac{\gamma_2^2 \lfloor \frac{p}{12} \rfloor^2}{2\sigma^2}\right) \exp\left(i\gamma_2 \xi \left\lfloor \frac{p}{12} \right\rfloor\right), \quad (13)$$

wherein $\gamma_2 \in \{0, \pm 1\}$. We define the *eigenprogression transform* of \mathbf{x} as the following rank-eight tensor:

$$\mathbf{U}_2(\mathbf{x})[t, p, q, \alpha_1, \beta_1, \alpha_2, \beta_2, \gamma_2] = \left| \mathbf{U}_1(\mathbf{x}) \overset{t, p, q}{*} \Psi_{(\alpha_2, \beta_2, \gamma_2)}^{\text{prog}} \right| [t, p, q]. \quad (14)$$

At first sight, Equation 10 suffers from an identifiability problem. Indeed, a different choice of basis for \mathbf{V}_k would incur a phase shift and/or a complex conjugation of the convolutional response $\mathbf{U}_1(\mathbf{x}) \overset{p, q}{*} \psi_k^{\text{Tonnetz}}$. Yet, because the eigenprogression transform consists of Gabor wavelets (i.e. with a symmetric amplitude profile) and is followed by a complex modulus operator, such differences in phase and/or spin are eventually canceled and thus have no effect on the outcome of the transform. Consequently, we pick one arbitrary pair $(\mathbf{v}_k^{\text{Re}}, \mathbf{v}_k^{\text{Im}})$ for each subspace \mathbf{V}_k , without loss of generality.

3. EXPERIMENTS

We evaluate the eigenprogression transform on a task of supervised composer recognition between Haydn and Mozart string quartets [7]. After averaging along time t and pitch p , we standardize each feature in the rank-five tensor

$$\mathbf{S}_2(\mathbf{x})[\alpha_1, \beta_1, \alpha_2, \beta_2, \gamma_2] = \sum_{t, p, q} \mathbf{U}_2(\mathbf{x})[t, p, q, \alpha_1, \beta_1, \alpha_2, \beta_2, \gamma_2] \quad (15)$$

to null mean and unit variance. Then, we train a linear support vector machine with $C = 10^4$, and report results with leave-one-out cross-validation. The ablation study in Table 1 confirm that all the five scattering variables ($\alpha_1, \beta_1, \gamma_1, \alpha_2, \beta_2, \gamma_2$) are beneficial to both sparsity and accuracy. However, because the dataset string quartet movements contains only 107 examples in total, the full eigenprogression transform (in dimension 8385) is exposed to statistical overfitting. For the sake of simplicity, rather than running a feature selection algorithm, we apply wavelet shrinkage denoising, i.e. we keep 1119 coefficients of largest energy on average, summing up to 50% of the total energy. This adaptive procedure has been proven to be near-optimal in the context of wavelet bases [3]. It leads to a state-of-the-art classification accuracy of 82.2%.

4. CONCLUSION

We have diagonalized the Laplacian of the Tonnetz graph and derived a multivariable scattering transform, named eigenprogression transform, that captures some local harmonic context in Western tonal music. Although the numerical example we gave was a task of composer recognition, the eigenprogression transform could, in principle, address other MIR tasks in the future, including cover song retrieval, key estimation, and structure analysis.

	dim.	ℓ_1/ℓ_2	acc. (%)
[7]			79.4
[8]			80.4
α_1	8	2.6	67.3
$\alpha_1 \beta_1$	24	4.6	71.0
$\alpha_1 \beta_1 \alpha_2$	129	6.1	72.0
$\alpha_1 \beta_1 \alpha_2 \beta_2$	1677	17.0	76.7
$\alpha_1 \beta_1 \alpha_2 \beta_2 \gamma_2$	8385	42.4	77.6
$\alpha_1 \beta_1 \alpha_2 \beta_2 \gamma_2$	1119	22.3	82.2

Table 1. Comparison between the eigenprogression transform and other transforms of smaller tensor rank, in terms of dimensionality, sparsity (ℓ_1/ℓ_2 ratio), and accuracy on a supervised task of composer recognition.

5. REFERENCES

- [1] Milton Babbitt. Twelve-tone invariants as compositional determinants. *The Musical Quarterly*, 46(2):246–259, 1960.
- [2] Rachel M. Bittner, Brian McFee, Justin Salamon, Peter Li, and Juan Pablo Bello. Deep salience representations for f_0 estimation in polyphonic music. In *Proc. ISMIR*, 2017.
- [3] David L Donoho and Jain M Johnstone. Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, 81(3):425–425, 1994.
- [4] Vincent Lostanlen and Stéphane Mallat. Wavelet scattering on the pitch spiral. In *Proc. DAFX*, 2015.
- [5] Stéphane Mallat. Group invariant scattering. *Commun. Pure Appl. Math.*, 65(10):1331–1398, 2012.
- [6] Oriol Nieto and Juan Pablo Bello. MSAF: Music structure analysis framework. In *Proc. ISMIR*, 2015.
- [7] Peter Van Kranenburg and Eric Backer. Musical style recognition: a quantitative approach. In *Handbook of Pattern Recognition and Computer Vision*, pages 583–600. World Scientific, 2005.
- [8] Gissel Velarde, Tillman Weyde, Carlos Eduardo Cancino Chacón, David Meredith, and Maarten Grachten. Composer recognition based on 2D-filtered piano-rolls. In *Proc. ISMIR*, 2016.