

Victor de Fontnouvelle  
12/10/2018

## Thesis Outline and Section

# Abstract

# Introduction

# Explanation of Methods

## Intuition

The Hodge Decomposition Theorem provides that  $\bar{Y}$  can be expressed as the sum of three orthogonal components:

1. The gradient flow  $G$
2. The curl flow  $F$
3. The harmonic flow  $H$

The gradient flow  $G$  represents a comparison matrix that corresponds directly to a vector  $v$  where each item is assigned a real value.  $G$  is the gradient of  $v$ , thus each entry is computed as  $G[ij] = v[j] - v[i]$ .

The harmonic flow  $H$  represents a ranking that is curl-free and divergence-free. This means that any three items in  $H$  will have pairwise rankings that are logically consistent. Specifically,  $H[ij] + H[jk] + H[kj] = 0, \forall i, j, k$ . Because  $\text{div}(H) = 0$ ,  $H$  contains plausible values in the sense that it could have been produced by real-world data. A harmonic flow thus indicates that there are cycles in the graph with more than three edges, where the edge weights don't sum to zero.

The curl flow  $C$  is the image of  $\text{curl}^*$ , the adjoint of the curl. Nonzero values of  $C$  thus indicate cycles of length three whose edge weights don't sum to zero.

The gradient flow provides the ranking that minimizes the least-squared residual, while the harmonic and curl flows characterize the residual. Specifically,  $\bar{Y} - H = G + C$ . A large curl flow indicates that the ordering of items ranked closely together is unreliable, while a large harmonic flow indicates that the ordering of items ranked further apart is unreliable. For example, if the curl flow is large but the harmonic flow is small, this indicates that the ranking is valid at a larger scale, but that the specific ordering of closely-ranked alternatives isn't very precise.

## Computation of $G$ , $C$ , and $H$

Gradient flow: We seek to find  $S \in \mathbb{R}^n$  that minimizes  $\|\delta_0 s - \bar{Y}\|$ . Such an  $s$  must satisfy  $\bar{Y} - \delta_0 s \in \delta_0^\perp = \ker(\delta_0^*)$ . Thus  $\delta_0^*(\bar{Y} - \delta_0 s) = 0 \implies \delta_0^* \delta_0 s = \delta_0^* \bar{Y}$ . Note that  $\Delta_0 = \delta_0^* \delta_0$  and  $\delta_0^* s = -\text{div}$ , thus  $s = -\delta_0^+ \text{div} \bar{Y}$ . We compute the gradient flow as  $G = \delta_0 s$ .

Curl flow: We seek a three-dimensional matrix  $M$  that minimizes  $\|\delta_1 * M - \bar{Y}\|$ . By similar logic,  $\text{curl} \text{curl}^* M = \text{curl} \bar{Y} \implies M = (\text{curl} \text{curl}^*)^{-1} \text{curl} \bar{Y}$ . Thus the curl flow  $C = \text{curl}^* (\text{curl} \text{curl}^*)^{-1} \text{curl} \bar{Y} = \text{curl}^+ \text{curl} \bar{Y}$ . We represent the curl matrix as a matrix with width  $\binom{n}{2}$  and height  $\binom{n}{3}$ , which transforms a vector holding all entries  $\bar{Y}[ij]$  s.t.  $j > i$  and outputs a vector corresponding to all triples  $(i, j, k)$  s.t.  $i < j < k$ . A triple  $(i, j, k)$  will be assigned the value  $\bar{Y}[ij] + \bar{Y}[jk] + \bar{Y}[ki]$ . For example, in the case of four variables, we have

$$\begin{bmatrix} 1 & -1 & 0 & 1 & 0 & 0 \\ 1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 1 & -1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & -1 & 1 \end{bmatrix} \begin{bmatrix} ab \\ ac \\ ad \\ bc \\ bd \\ cd \end{bmatrix} = \begin{bmatrix} abc \\ abd \\ acd \\ bcd \end{bmatrix}$$

To check correctness, we examine the triple  $(a, b, c)$ . This triple is assigned the value  $\bar{Y}[ab] - \bar{Y}[ac] + \bar{Y}[bc]$ . This equals  $\bar{Y}[ac] + \bar{Y}[bc] + \bar{Y}[ca]$ , as  $-\bar{Y}[ac] = \bar{Y}[ca]$  since  $\bar{Y}$  is symmetric. This is the curl, as we expect.

Harmonic flow: The Hodge decomposition theorem provides that  $\bar{Y} = G + H + C$ , thus we compute  $H$  as  $\bar{Y} - G - C$ .

## Application to Dataset

## Conclusion