

MINIMAL UNIMODAL DECOMPOSITION ON TREES

YULIY BARYSHNIKOV AND ROBERT GHRIST

ABSTRACT. The decomposition of a density function on a domain into a minimal sum of unimodal components is a fundamental problem in statistics, leading to the topological invariant of *unimodal category* of a density. This paper gives an efficient algorithm for the construction of a minimal unimodal decomposition of a tame density function on a finite metric tree.

1. INTRODUCTION

1.1. **Motivation.** In data analysis, the operation of clustering is fundamental. At its base is a problem wedged between geometry and topology: given a set of points and a notion of distance or proximity among them, compute a parsimonious division into sets of points that are mutually *close*. In practice, this is a subtle question that depends sensitively on the models of proximity for the input and the desired nature of the output [13, 3].

The discretized nature of this problem obscures the more topological (and ancient) problem of minimal decomposition of a space into simple pieces or, suggestively, *modal domains*. For a geometric domain $D \subset \mathbb{E}^n$, one might ask for the minimal number of convex pieces into which D can be decomposed, minimal number providing an unambiguous descriptor of geometric complexity. For a topological space X , the minimal mode is not a convex domain but rather a *contractible* subset $U \subset X$, meaning that U has the homotopy type of a point. The minimal number of such subsets covering X is called the *geometric category* of X and is a homeomorphism invariant of X . There is a family of related topological categories, the most important being the *Lusternik-Schnirelmann category* (using subsets nullhomotopic in X) which branches quickly into homotopy theory [5]. Other related notions of minimal topological decomposition include the classical *sectional category* (or *Schwarz genus*) of a fibration and the more modern variant of *topological complexity* of path planning [7].

This paper concerns a weighted version of these problems adapted to statistics. Given a *density* $f: X \rightarrow [0, \infty)$, what is the minimal number of *modes* into which it can be decomposed as a sum? In the geometric version of this problem, one natural notion of a mode is a Gaussian, and the problem of minimal approximation of a density as a sum of Gaussians is well-studied on one-dimensional domains [2, 6, 14]. This choice of a Gaussian as fundamental mode is somewhat geometrically rigid, analogous to the clustering of a space into convex pieces. One can imagine other basis unimodal distributions [11, 12].

This paper concerns the coordinate-free topological version of the problem of minimal decompositions of densities over a space — a statistical analogue of the category of a domain. The resulting invariant is called the *unimodal category* [1, 8].

1.2. **Unimodal category.** For X a topological space, let $\mathfrak{D} = \mathfrak{D}(X)$ denote the set of all compactly supported continuous functions $f: X \rightarrow [0, \infty)$. Such a function $u \in \mathfrak{D}$ is *unimodal* if the upper excursion sets $u^c = u^{-1}([c, \infty))$ have the homotopy type of a point for all $0 < c \leq M$ and are empty for all $c > M$. Such a u has M as its maximal value.

Work supported by the National Science Foundation via DMS-1622370.

Work supported by the Office of the Assistant Secretary of Defense Research & Engineering through ONR N00014-16-1-2010.

We will refer to the nonempty upper excursion sets $u^c \subset X$ as being *contractible*, though it must be clarified that such sets are contractible *in themselves* as opposed to being merely contractible in X . The latter would be more in line with the definitions used in Lusternik-Schnirelmann theory, but is less relevant for most applications (where $X = \mathbb{R}^n$).

Definition 1.1. *The unimodal category of $f \in \mathcal{D}(X)$ is the minimal number $\text{ucat}(f)$ of unimodal distributions $u_\alpha, \alpha = 1, \dots, \text{ucat}(f)$ on X such that f is the pointwise sum of the collection $\{u_\alpha\}$.*

In the data analysis interpretation of the unimodal functions, where the mode corresponds to *signal*, and the spread of the density around it, to *noise*, it makes sense to assume some similarities of the noise generating mechanisms for different modes. In the world of Gaussian distribution this leads to assumption of the fixed, or slowly varying covariant form. A parsimonious, homeomorphism-invariant version would assume a much weaker formulation, which nonetheless strengthens significantly the notion of unimodal category, as follows:

Definition 1.2. *The strong unimodal category of $f \in \mathcal{D}(X)$ is the minimal number $\text{ucat}(f)$ of distributions $u_\alpha, \alpha = 1, \dots, \text{ucat}(f)$ on X that sum up to f such that any intersection of the upper excursion sets is either empty or contractible.*

Since unimodal functions remain unimodal under a homeomorphic change of coordinates, the unimodal category is a coordinate-free invariant of a density. This makes it of significant potential use in applications where data is collected from noisy or otherwise uncertainly located samples. The initial paper on the subject gave a constructive algorithm for computing the unimodal category on \mathbb{R}^1 , along with generalizations to unimodal categories based on pointwise norms rather than addition [1]. The thesis of Govc showed just how subtle and difficult the problem of computation of these invariants is in higher dimensions [9]. Recent applications of the unimodal category in 1-d by Huntsman are currently being applied to problems of mixture estimation in statistics [10].

We believe that *strong* unimodal category would prove more amenable to analysis, but postpone this line of research till later, noting only that for our model, where the underlying topological space is a metric tree, the notion of strong unimodal category is identical with the standard one.

1.3. Contribution. The primary contribution of this note is the presentation and proof of correctness of an efficient greedy algorithm for the computation of a minimal unimodal decomposition of a density over a tree, this yielding the unimodal category. More than simply computing a topological invariant, this method permits identification of “essential” modes which are local maxima for *any* minimal unimodal decomposition. We believe that this extension will permit novel applications of the unimodal category and point the way to computational methods on suitably restricted higher dimensional domains.

As a perhaps fanciful toy problem for the sake of motivation, consider the following scenario. Suppose one wishes to detect radiological substances by means of crude sensors mounted on vehicles in traffic on city streets. At any given time, the network of vehicles returns a sampling of a distribution on a graph (radioactivity levels restricted to the idealized 1-d cell complex of city streets). It is perhaps known that there are certain expected modes — at, say, hospitals, research facilities, and universities. In a dense urban environment, one might have difficulty distinguishing true modes from interference modes, and the problem of noise in position measurement is a further complicating feature. It is in such circumstances that a minimal unimodal decomposition may assist as an unambiguous warning of a significant unexpected mode.

2. SETUP

2.1. Linearization. Let Γ be a finite metric tree, that is, a 1-dimensional compact contractible metric space stratified into 0-dimensional *vertices* and 1-dimensional (open) *edges*. All subtrees of Γ are assumed to inherit both the cell structure and the metric of Γ . The subtrees of Γ form a finite lattice.

In order to work with finite unimodal decompositions, we henceforth assume that $\mathcal{D}(\Gamma)$ is restricted to functions with a finite number of critical points. Given such an $f \in \mathcal{D}(\Gamma)$, we can assume

that the restriction of the function to each edge is weakly monotonic, by adding new vertices at local maxima and minima. Using the restriction of f to each edge (or an affine linear function of it) as a coordinate, we can also assume that f is *edge-linear*, that is, f restricted to each edge is an affine function of that edge with respect to the internal metric structure: see Figure 1.

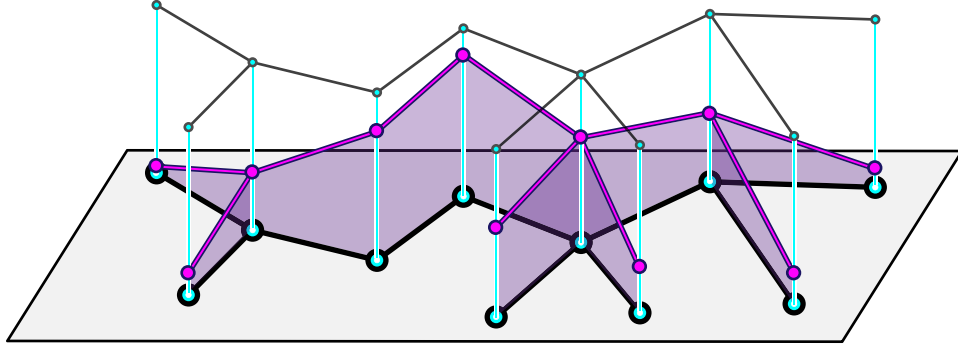


FIGURE 1. Example of an edge-linear unimodal function on a tree.

Lemma 2.1. *If f is constant on an edge e of Γ , then the operation of contraction by e taking Γ to the tree Γ/e sends unimodal decompositions of f on Γ to unimodal decompositions of f on Γ/e preserving number of modes. This operation is reversible, taking any unimodal decomposition of f on Γ/e to a mode-preserving decomposition of f on Γ .*

Lemma 2.2. *If $f = \sum f_\alpha$ is a unimodal decomposition of an edge-linear $f \in \mathfrak{D}(\Gamma)$, then its components f_α can be chosen to be edge-linear as well.*

Proof. Replacing a unimodal component f_α with f_α^l with its edge-linear interpolation, preserving the values at the vertices of Γ , is again unimodal, and that the resulting components sum up to f , as their values at vertices do. \square

As an immediate corollary we obtain the following.

Lemma 2.3. *Any unimodal decomposition can be modified (without changing the number of components) so that all component modes are maximized at the vertices of Γ .*

As a result of Lemmas 2.1-2.3, we henceforth assume that all $f \in \mathfrak{D}(\Gamma)$ and all modes in unimodal decompositions thereof are edge-linear with modes at vertices.

2.2. Free and forced. Given the data $f \in \mathfrak{D}(\Gamma)$, we call a subtree $\Gamma' \subset \Gamma$ *mode-free* if there exists a unimodal decomposition of f such that Γ' is free of modes. Clearly, mode-free subtrees are closed under the operation of taking subtrees. We will call a vertex of Γ *mode-forced* for f if v is a mode of every minimal unimodal decomposition of f on Γ .

Lemma 2.4. *For any $f \in \mathfrak{D}(\Gamma)$ satisfying the assumptions of §2, a mode-forced vertex exists.*

Proof. Choose any vertex v and consider it the root of Γ ; then Γ becomes a union of several branches each having v as a root. We will call a branch (a connected component of the complement of v , and all vertices in it excluding v) *insignificant*, if the values of f are monotonically decreasing away from v in that branch.

It is clear that removing all insignificant vertices does not increase the unimodal category of f (but can generate more insignificant vertices). Pruning iteratively the tree of its insignificant vertices results either in a single vertex (in which case the original function f is unimodal), or in a tree with at least two leaves, each of which is mode-forced. \square

Specific examples of mode-forced vertices include the global maximum of f , as well as all local maxima of Γ such that all but one of the components of their complement are monotonically decreasing paths to leaves.

3. SWEEPING

3.1. Sweeping operation. In this section we define a *sweeping operation* that will generate a function with a given mode.

Let v be a vertex of Γ and $f \in \mathcal{D}(\Gamma)$ satisfying the assumptions of §2.1. Define the function $h_{f,v}$ on Γ using the following procedure.

- (1) Orient all edges of Γ away from v (making it the root of the oriented tree).
- (2) Set $h_{f,v}(v) = f(v)$.
- (3) If for an oriented edge, $u \rightarrow w$, $h_{f,v}(u)$ is defined, then apply the *sweeping move*:

$$(1) \quad h_{f,v}(w) = \begin{cases} h_{f,v}(u) & \text{if } f(u) < f(w); \\ \max(h_{f,v}(u) - (f(u) - f(w)), 0) & \text{otherwise} \end{cases}$$

Figure 2 below illustrates the sweeping mechanism.

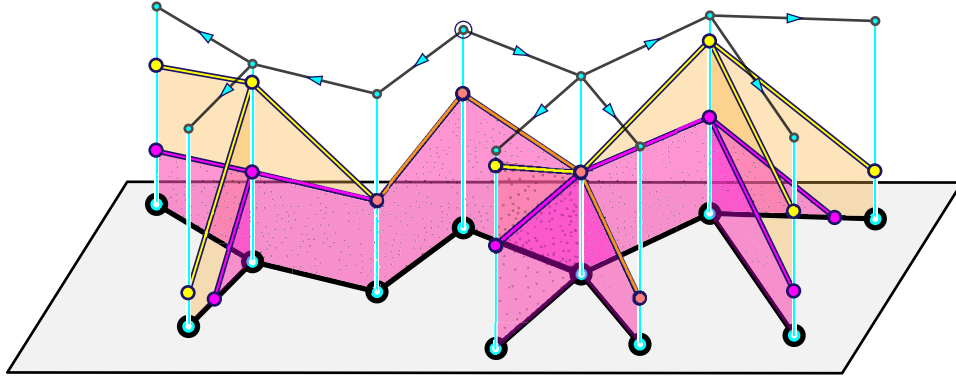


FIGURE 2. Propagating $h_{f,v}$ (pink) given f (yellow) from v (the central mode). Note that the sweeping process can create new interior nodes in the trees where the function hits zero.

3.2. Remainders and freedom. Given a vertex $v \in \Gamma$, we will call the function $R_v f = f - h_{f,v}$ the v -remainder. This is nonnegative as, clearly, $h_{f,v} \leq f$ for any $v \in \Gamma$. Also noted is that the support of $R_v f$ does not contain v . The following is the critical result needed for our constructions.

Theorem 3.1. *If $\Gamma' \subset \Gamma$ is a mode-free subtree for f , and $v \notin \Gamma'$, then Γ' is also mode-free for $R_v f$.*

In other words, taking v -remainders does not shrink free subtrees not containing v . We postpone the proof until §5.2.

4. GREEDY ALGORITHM

The algorithm given in [1] for computing a minimal unimodal decomposition of $f \in \mathcal{D}([a, b])$ involved sweeping from a to b , identifying mode-forced vertices, then removing their contribution by computing remainders, until the entire interval was swept.¹ In essence, the same process is here employed for a tree in the following *greedy algorithm*:

- (1) Using Lemma 2.4, find a mode-forced vertex v .

¹That is, using the language of this paper.

The following lemma asserts that we can restrict our attention to admissible modifications of aggregations of leaf functions.

Lemma 5.2. *Consider an edge, $[uw]$ and a collection of leaf functions $f_\lambda, \lambda \in \Lambda$, each non-increasing from u to w . Their sum $f_\Lambda = \sum_{\lambda \in \Lambda} f_\lambda$ is also non-increasing in the same direction. Let g_Λ be an admissible modification of f_Λ . Then there exist admissible modifications g_λ of each $f_\lambda, \lambda \in \Lambda$ so that $\sum_{\lambda \in \Lambda} g_\lambda = g_\Lambda$ on $[u, w]$.*

Proof. By admissibility, we have that $f_\Lambda(u) \geq g_\Lambda(u) \geq g_\Lambda(w) \geq f_\Lambda(w)$. The claim is equivalent to existence of a solution to the following systems of (in)equalities on the variables $g_\lambda(\cdot)$:

$$(3) \quad \begin{aligned} f_\lambda(u) &\geq g_\lambda(u), & \lambda \in \Lambda \\ g_\lambda(u) &\geq g_\lambda(w), & \lambda \in \Lambda \\ g_\lambda(w) &\geq f_\lambda(w), & \lambda \in \Lambda \\ \sum_{\lambda \in \Lambda} g_\lambda(u) &= g_\Lambda(u), \\ \sum_{\lambda \in \Lambda} g_\lambda(w) &= g_\Lambda(w). \end{aligned}$$

Farkas' lemma implies that the solution to (3) *does not* exist if and only if the dual system (in variables $\mu_\lambda^u \geq 0, \mu_\lambda \geq 0, \mu_\lambda^w \geq 0, \nu_u, \nu_w$), corresponding to the lines in (3)) has a solution:

$$(4) \quad -\mu_\lambda^u + \mu_\lambda - \nu_u = 0;$$

$$(5) \quad \mu_\lambda^w - \mu_\lambda + \nu_w = 0;$$

$$(6) \quad \sum_{\lambda} ((-\mu_\lambda^u f_\lambda(u) + \mu_\lambda^w f_\lambda(w)) - \nu_u g_\Lambda(u) + \nu_w g_\Lambda(w)) > 0.$$

Using the fact that $\sum_{\lambda} f_\lambda(u) = f_\Lambda(u)$ and $\sum_{\lambda} f_\lambda(w) = f_\Lambda(w)$, transform the left hand side of (6) as

$$(7) \quad 0 < \sum_{\lambda} ((-\mu_\lambda^u f_\lambda(u) + \mu_\lambda^w f_\lambda(w)) - \nu_u g_\Lambda(u) + \nu_w g_\Lambda(w)) =$$

$$(8) \quad \sum_{\lambda} ((-\mu_\lambda + \nu_u) f_\lambda(u) + (\mu_\lambda - \nu_w) f_\lambda(w)) - \nu_u g_\Lambda(u) + \nu_w g_\Lambda(w) =$$

$$(9) \quad \sum_{\lambda} \mu_\lambda (-f_\lambda(u) + f_\lambda(w)) + \nu_u (f_\Lambda(u) - g_\Lambda(u)) + \nu_w (g_\Lambda(w) - f_\Lambda(w)).$$

Further, (5) implies that $\nu_u \leq \mu_\lambda$ for all λ , and therefore $\nu_u \leq \min \mu_\lambda =: \mu$. Similar reasoning (using (6)) leads also to $\nu_w \leq \mu$.

Using these inequalities with (9), and taking into account that $f_\Lambda(u) \geq g_\Lambda(u) \geq g_\Lambda(w) \geq f_\Lambda(w)$, we conclude that (9) is bounded above by

$$(10) \quad \mu(-f_\Lambda(u) + f_\Lambda(w)) + \mu(f_\Lambda(u) - g_\Lambda(u)) + \mu(g_\Lambda(w) - f_\Lambda(w)) \leq 0,$$

contradicting the implication (7) of non-existence of a solution to (3). \square

6. STABILITY OF MODE-FREE SUBTREES

In this section we finally prove Theorem 3.1 on the stability of mode-free subtrees.

Proof of Theorem 3.1. Lemmas 5.1 and 5.2 imply that any admissible modification of f on an edge of Γ' maintains a leaf function decomposition, and thus, the mode-free nature of Γ' . To show that $R_v f$ preserves the mode-free subtree Γ' for $v \notin \Gamma'$, we apply the sweeping algorithm generating $h_{f,v}$ along the edges in Γ' , propagating away from λ^* (the leaf of Γ' adjacent to the component of $\Gamma - \Gamma'$ containing v) and performing admissible modifications.

Specifically, if we modify the leaf functions so that $f_{\lambda^*} = h_{f,v} + C$ for some constant $C \geq 0$, then we have a leaf function decomposition of $R_v f$ on Γ' . If $f_{\lambda^*}(\lambda^*) < h_{f,v}(\lambda^*)$, then raising $f_{\lambda^*}(\lambda^*)$ to $h_{f,v}(\lambda^*)$ and lowering the other leaf functions is admissible. In this case, $C = 0$.

Fix an edge $[uw]$ where the leaf functions are to be modified. We assume that the propagation happens from u to w (that is λ^* is in the component of $\Gamma - (u, w)$ containing u). Consider the values $f_u = f(u)$ and $f_w = f(w)$ at the vertices. Decompose these as follows:

$$(11) \quad f_u = f_u^* + f_u^{\rightarrow} + f_u^{\leftarrow} \quad f_w = f_w^* + f_w^{\rightarrow} + f_w^{\leftarrow}$$

where $f_u^* = f_{\lambda^*}(u)$ and $f_w^* = f_{\lambda^*}(w)$ are values of the leaf function corresponding to λ^* ; the f_u^{\rightarrow} , f_w^{\rightarrow} , f_u^{\leftarrow} , and f_w^{\leftarrow} terms are values (at u and w) of sums of the remaining leaf functions that are respectively oriented (nonincreasing) from u to w (\rightarrow) and from w to u (\leftarrow). By Lemma 5.2, it suffices to give admissible modifications of these aggregated leaf functions — that is of values f_u^{\rightarrow} , f_w^{\rightarrow} , f_u^{\leftarrow} , and f_w^{\leftarrow} .

After potentially subdividing Γ based on where $h_{f,v}$ hits zero, there are two cases, corresponding to f being increasing or decreasing from u to v . If f is decreasing, then $f_u - f_w \leq f_u^* - f_w^*$, so that its decrement is less or equal than the decrement of f_{λ^*} . For these cases, we modify f -values of leaf functions at u and w to g -values as per Lemma 5.2. These modifications are given in Table 6 and can be checked as admissible and summing to f as per (11).

CASE	g_u^{\rightarrow}	g_u^{\leftarrow}	g_u^*	g_w^{\rightarrow}	g_w^{\leftarrow}	g_w^*
$f_u - f_w \geq 0$	f_u^{\rightarrow}	f_u^{\leftarrow}	f_u^*	f_u^{\rightarrow}	$f_w - f_u + f_u^{\leftarrow}$	f_u^*
$f_u - f_w < 0$	f_u^{\rightarrow}	f_u^{\leftarrow}	f_u^*	f_u^{\rightarrow}	f_u^{\leftarrow}	$f_u^* - f_u + f_w$

Sweeping over Γ' , this modifies the leaf function at λ^* to be of the form $h_{f,v} + C$. Thus, modifying f to the remainder $R_v f$ has the effect of maintaining a leaf function decomposition on Γ' . \square

Corollary 6.1. *Under the assumptions on $f \in \mathcal{D}(\Gamma)$ from §2.1, the greedy algorithm of §4 returns a minimal unimodal decomposition.*

7. CONCLUSIONS

The present work gives a constructive method for the computation of minimal unimodal decompositions of densities on trees. We end with a few remarks.

- (1) Unimodal decompositions are far from unique. Minimal unimodal decompositions are likewise not unique, but in a structured manner. The mode-forced vertices and mode-free subtrees provide a skeleton on which a given density hangs. By performing the sweeping moves of the greedy algorithm in different “directions” one arrives at many minimal unimodal decompositions. The analogous variations over an interval consist in sweeping from left-to-right or right-to-left [1].
- (2) Assume that the values of f on the V vertices of Γ are given as the input. As written, the greedy algorithm for generating a unimodal decomposition is $O(V^2)$. One can do better. For example, as $V \rightarrow \infty$, it is reasonable to assume that modes have a uniformly bounded support, in which case the sweeping is absorbed into the constants, and the algorithm drops to $O(\text{ucat}(f))$.
- (3) Of course, the restriction of these results to trees is suboptimal. For applications of unimodal decompositions in disciplines where precise geometric data can be hard to come by (e.g., phylogenetics or neuroscience), graphs with cycles are not only possible but (especially in the case of neuroscience) critical. The generalization from graphs to higher dimensional domains is likewise important but appears formidable, depending on the model of unimodal decomposition employed. We view the construction of a minimal unimodal decompositions on graphs to be an important prerequisite for this challenge.

REFERENCES

- [1] Y. Baryshnikov and R. Ghrist, “Unimodal Category and Topological Statistics,” *Proc. NOLTA*, 2011.
- [2] J. Behboodan, “On the modes of a mixture of two normal distributions,” *Technometrics*, 12, 1970, 131–139.
- [3] G. Carlsson & F. Mémoli, “Classifying clustering schemes,” *Found. Comput. Math.*, 13 (2):221–252 (2013).

- [4] M. Carreira-Perpinan and C. Williams, "On the number of modes of a Gaussian mixture," in *Scale-Space Methods in Computer Vision, Lecture Notes in Comput. Sci.*, vol. 2695, 2003, 625–640.
- [5] O. Cornea, G. Lupton, J. Oprea, and D. Tanré, *Lusternik-Schnirelmann Category*, Amer. Math. Soc., 2003.
- [6] I. Eisenberger, "Genesis of bimodal distributions," *Technometrics*, 6, 1964, 357–363.
- [7] M. Farber, "Topological complexity of motion planning," *Discrete Comput. Geom.*, 29 (2), 2003, 211–221.
- [8] R. Ghrist, *Elementary Applied Topology*, Createspace, 2014.
- [9] D. Govc, "Unimodal category and the monotonicity conjecture," arXiv:1709.06547 [math.AT], 2017.
- [10] S. Huntsman, "Topological Mixture Estimation," preprint, 2018.
- [11] I. Kakiuchi, "Unimodality conditions of the distribution of a mixture of two distributions," *Math. Sem. Notes Kobe Univ.*, 9, 1981, 315–325.
- [12] J. Kemperman, "Mixtures with a limited number of modal intervals," *Ann. Statist.* 19, 1991, 2120–2144.
- [13] J. Kleinberg, "An impossibility theorem for clustering," in *Proc. NIPS*, 2002, 446–453.
- [14] C. Robertson and J. Fryer, "Some descriptive properties of normal mixtures," *Skand. Aktuarietidskr*, 1969, 137–146.

DEPARTMENTS OF MATHEMATICS AND ELECTRICAL & COMPUTER ENGINEERING, UNIVERSITY OF ILLINOIS
E-mail address: ymb@uillinois.edu

DEPARTMENTS OF MATHEMATICS AND ELECTRICAL & SYSTEMS ENGINEERING, UNIVERSITY OF PENNSYLVANIA
E-mail address: ghrist@math.upenn.edu