DIAGONALIZING THE GENOME I

SATYAN L. DEVADOSS AND JACK MORAVA

Abstract. We construct an equivariant blowup of the space of real symmetric matrices, with a natural stratification indexed by trees, which keep track of eigenvalue coincidences. We suggest, without much justification, that systems of oscillators parametrized by these spaces may provide useful models in genomics.

1. Introduction

The space $\mathcal{Q}_n$ of real symmetric $n \times n$ matrices (equivalently, of self-adjoint operators, or quadratic forms in n variables) parametrizes, among other things, systems of coupled harmonic oscillators, and is thus of fundamental importance in mathematical mechanics. Nearly forty years ago, V.I. Arnol'd observed [4] that this (contractible) space has a very interesting stratification defined by eigenvalue multiplicities, and he remarked its relevance in applications to phenomena involving resonance [26]. In this paper we construct an $(\mathbb{S}_n \times \mathrm{SO}_n)$-equivariant blowup

$$\begin{array}{ccc} \widetilde{\mathcal{Q}_n} & \dashrightarrow & \mathcal{Q}_n \\ \downarrow & & \downarrow \\ \overline{\mathcal{M}}^{\mathrm{or}}_{0,n+1}(\mathbb{R}) & \longrightarrow & \mathrm{SP}^n(\mathbb{R})/\mathrm{Aff}_+ \end{array}$$

of $\mathcal{Q}_n$ (modulo affine equivalence), which, roughly speaking, resolves these eigenvalue coalescences. The space $\widetilde{\mathcal{Q}_n}$ has a stratification generalizing Arnol'd's, replacing his eigenvalue partitions with trees having those partitions as collections of leaves. Final remarks in Section 6 suggest some applications to theoretical biology, in particular, to models for the relation of genotype to phenotype.

The first section below summarizes some familiar facts about quadratic forms, and the following three sections contain the new material of the paper: they are an introduction to certain remarkable compact hyperbolic manifolds $\overline{\mathcal{M}}_{0,n}(\mathbb{R})$ and their orientation covers $\overline{\mathcal{M}}^{\mathrm{or}}_{0,n}(\mathbb{R})$. The former spaces first appeared as the real points of moduli spaces of stable genus zero algebraic curves marked with families of distinct smooth points (which can also be interpreted as compactifications of the moduli spaces of ordered configurations

2000 *Mathematics Subject Classification.* 14H10, 92B10, 16E50.
Key words and phrases. phylogenetics, configuration spaces, real symmetric matrices, genome.

of points on the real projective line), but here we understand them as spaces of rooted metric trees with labeled leaves, as in Figure 1. Among other things, they resolve the singularities of the spaces of metric trees studied, from the phylogenetic point of view, by Billera, Homes, and Vogtmann [6].

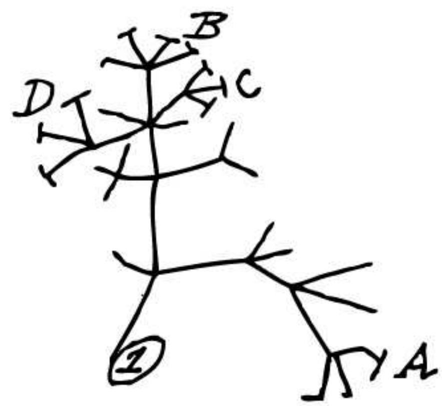

FIGURE 1. An example from Darwin's notebook (1837); see section 6.2.

One of the objectives of this paper is to provide these tree spaces with natural pseudometrics and measures. These spaces have some remarkable similarities [20] to the space forms which captured the imagination of geometers in the nineteenth century: they are $K(\pi, 1)$ manifolds with very pretty tesselations. The work of Klein and others soon found important applications in physics and mechanics, and applications of our generalized space forms may only now be appearing on the horizon.

Acknowledgments. We would like to thank Charles Epstein for suggesting that speciation might profitably be regarded as a kind of resonance [16], Amnon Ne'eman for advice on real algebraic geometry, Sikimeti Mau for insights into configuration spaces, and especially Jim Stasheff for continued interest and encouragement over many years. SD also thanks Lior Pachter and the University of California, Berkeley, and Clemens Berger and the Université de Nice for their hospitality during his sabbatical visits. JM was partially supported by the DARPA FunBio program, and none of it would have ever been possible without Ben Mann's visionary work spearheading that project.

2. Quadratic forms and their eigenvalues

2.1. The geometry of spaces of matrices with conditions on their eigenvalues is complicated. The orthogonal group SO_n of isometries of Euclidean space $\mathbb{R}^n$ acts almost freely on $\mathcal{Q}_n$ by conjugation, and quadratic forms are classified up to isomorphism by elements of the resulting quotient. Since

$$\dim \mathcal{Q}_n - \dim \mathrm{SO}_n \;=\; \binom{n+1}{2} - \binom{n}{2} \;=\; n\;,$$

it is natural to think of an equivalence class as indexed by its configuration

$$\mathcal{Q}_n \;\to\; \mathrm{SP}^n(\mathbb{R}) \;:\; Q \;\mapsto\; \boldsymbol{\lambda}(Q) := \{\lambda_1 \leq \cdots \leq \lambda_n\}$$

of eigenvalues in the symmetric product $\mathrm{SP}^n(\mathbb{R}) := \mathbb{R}^n/\mathbb{S}_n$. When these eigenvalues are all distinct, the numbers

$$\delta_k \;:=\; \frac{\lambda_{k+1}-\lambda_k}{\lambda_n-\lambda_1} > 0 \qquad\qquad \text{for } k \in \{1,\ldots,n-1\}$$

are well-defined and sum to 1, defining an element $\boldsymbol{\delta}(Q) := (\delta_1,\ldots,\delta_{n-1})$ in the interior of the $(n-2)$-simplex

$$\Delta^{n-2} \;:=\; \{(x_1,\ldots,x_{n-1}) \in \mathbb{R}^{n-1}_{\geq 0} \;|\; \sum x_i = 1\}\;.$$

The contractible group

$$\mathrm{Aff}_+ \;:=\; \{x \mapsto ax+b \;|\; a \in \mathbb{R}^\times_+,\; b \in \mathbb{R}\}$$

of affine transformations of $\mathbb{R}$ acts freely on $\mathcal{Q}_n$, and the quotient of $\mathcal{Q}_n$ by SO_n is Aff_+-equivariantly isomorphic, by the map

$$\mathcal{Q}_n/\mathrm{SO}_n \to \mathbb{R}^2 \times \Delta^{n-2} \;:\; \boldsymbol{\lambda}(Q) \mapsto (\lambda_n-\lambda_1,\lambda_1) \times \boldsymbol{\delta}(Q)\,,$$

to a product of the plane with a simplex. We can therefore interpret the quotient $\mathcal{Q}_n/\mathrm{Aff}_+ \times \mathrm{SO}_n$ as a space of quadratic forms normalized by suitable conditions on the trace and determinant.

2.2. The configuration space of n ordered, labeled distinct points on the real line can be defined as

$$\mathrm{Config}^n(\mathbb{R}) \;:=\; \mathbb{R}^n - \{(x_1,\ldots,x_n) \in \mathbb{R}^n \;|\; \exists\, i \neq j,\; x_i = x_j\}\;.$$

Any n-tuple $(v_1,\ldots,v_n) \in \mathbb{R}^n$ of distinct real numbers has an intrinsic order

$$v_{\sigma(1)} \;<\; \cdots \;<\; v_{\sigma(n)}$$

which defines an element σ of the symmetric group $\mathbb{S}_n$ of permutations of n things, and the construction in the preceding paragraph extends to an $(\mathrm{Aff}_+ \times \mathbb{S}_n)$-equivariant homeomorphism

$$\mathrm{Config}^n(\mathbb{R}) \;\simeq\; \mathbb{R}^2 \times \overset{o}{\Delta}{}^{n-2} \times \mathbb{S}_n\;.$$

The $n!$-sheeted ramified cover

$$\mathbb{R}^n \;\to\; \mathrm{SP}^n(\mathbb{R}) \;\simeq\; \mathbb{R}^n \;:\; (v_1,\ldots,v_n) \;\mapsto\; (e_1,\ldots,e_n)$$

defined by the elementary symmetric functions

$$\prod (t - v_k) \; = \; \sum (-1)^k e_k t^k$$

is equivariant with respect to the action $v_k \mapsto av_k + b$ of the affine group on the left, provided it acts by

$$e_k \mapsto \sum_{k \geq l \geq 0} \binom{n-l}{n-k} a^l b^{k-l} e_l$$

on the right. The diagram

$$\Delta^{n-2} \times \mathrm{Aff}_+ \xleftarrow{\simeq} \mathrm{SP}^n(\mathbb{R}) \xrightarrow{v \mapsto e} \mathbb{R}^n$$

then defines an Aff_+-equivariant map

$$(\Delta^{n-2}, \partial\Delta^{n-2}) \times \mathrm{Aff}_+ \; \to \; (\mathbb{R}^n, \mathrm{Disc}_n)$$

where the discriminant locus on the right is the inverse image of zero under

$$(e_k) \; \mapsto \; \prod_{i \neq k} (v_i - v_k)^2 \; .$$

As the eigenvalues of a matrix coalesce, its image in $(\mathcal{Q}_n/\mathrm{Aff}_+) \times \mathrm{SO}_n$ approaches the boundary of Δ^{n-2}, which lies over the classical discriminant locus. The pullback

$$\begin{array}{ccc} \widetilde{\mathcal{Q}^0_n} & \xrightarrow{p_1} & \mathrm{Config}^n(\mathbb{R}) \\ {\scriptstyle p_0}\downarrow & & \downarrow{\scriptstyle /\mathbb{S}_n} \\ \mathcal{Q}^0_n & \xrightarrow{/\mathrm{SO}_n} & \mathbb{R}^n - \mathrm{Disc}_n \end{array}$$

defines a cover (with $n!$ components) of the space $\mathcal{Q}^0_n$ of quadratic forms with distinct eigenvalues. Because $\widetilde{\mathcal{Q}_n}$ is defined as a fiber product, there is a canonical $(\mathbb{S}_n \times \mathrm{SO}_n \times \mathrm{Aff}_+)$-equivariant map

$$\widetilde{\mathcal{Q}^0_n} \to \widetilde{\mathcal{Q}_n} \to \mathcal{Q}_n$$

unramified away from the subspace of matrices with coincident eigenvalues. And because both $\mathcal{Q}_n$ and Δ^{n-2} are contractible, $\widetilde{\mathcal{Q}_n}$ is aspherical, with fundamental group isomorphic to that of $\overline{\mathcal{M}}^{\mathrm{or}}_{0,n+1}(\mathbb{R})$ (defined in the following sections).

3. Navigation in tree-space

3.1. A *metric tree* is a tree with a nonnegative weight assigned to each of its *internal* edges [5]. Billera, Holmes, and Vogtmann [6] have constructed a space BHV_n of isometry classes of *rooted* metric trees with n labeled leaves. When such a tree is binary, it has $n - 2$ internal nodes, and hence $2n - 2$ nodes in all. Diaconis and Holmes [15] have shown that such trees, with all nodes labelled, are in bijection with the $(2n-3)!!$ possible (unordered) pairings of those nodes. This enumeration defines coordinate patches for the space of such trees, parametrized by the weights of their internal edges. The resulting space is contractible (in fact a cone), formed by gluing $(2n-3)!!$ orthants $\mathbb{R}^{n-2}_{\geq 0}$, one for each type of labeled binary tree. As the weights go to infinity, we get degenerate trees on the boundaries of the orthants. Two boundary faces are identified when they contain the same degenerate trees.

Boardman's subspace $\mathcal{T}_n \subset \mathrm{BHV}_n$ of *fully-grown* rooted trees [7] consists of (equivalence classes of) trees with weights in $[0, 1]$, having at least one edge of weight one. If a tree is not a corolla (with all internal weights zero), then the tree can be rescaled (by dividing its weights by their maximum) to be full-grown. This defines a homeomorphism with the (unreduced) smash product

$$\mathrm{BHV}_n \;\simeq\; (\mathbb{R}_{\geq 0}, 0) \wedge \mathcal{T}_n$$

which sends a tree with edges of weight $\{w_k\}$ to the pair defined by $\max\{w_k\} \in \mathbb{R}_+$, together with a full-grown rescaled version of that tree; the corolla is sent to the cone point $0 \times \{\mathcal{T}_n\}$.

Theorem 1. *The one-point compactification* BHV_n^+ *is homeomorphic to the (unreduced) suspension of* $\mathcal{T}_n$*.*

Figure 2 displays $\mathcal{T}_4$, seen as the Peterson graph, with 10 vertices corresponding to rooted trees with four leaves with one internal edge of weight 1; the 15 edges belong to trees with two internal weights summing to 1.

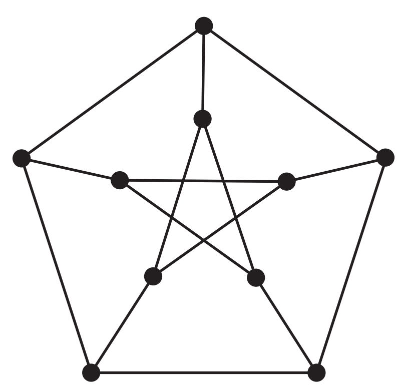

FIGURE 2. The space $\mathcal{T}_4$ as the Peterson graph.

The subspace of BHV_n of trees with weights that sum to one is also homeomorphic to $\mathcal{T}_n$, by rescaling the maximal edge weight to one. $\mathcal{T}_n$ also appears in [33] as the *tropical Grassmannian* $\mathcal{G}'''_{2,n+1}$. These spaces of trees have an important literature in representation theory and combinatorics [32]. The symmetric group $\mathbb{S}_n$ acts on $\mathcal{T}_n$ by permuting the labels of the leaves; this restricts to an $\mathbb{S}_{n-1}$-equivariant homotopy equivalence

$$\mathcal{T}_n \sim S^{n-3} \wedge \mathbb{S}_{n-1}^+ .$$

Corollary 2. *The space* BHV_n^+ *is homotopy-equivalent to a wedge of* $(n-1)!$ *spheres of dimension* $n-2$.

In fact, by dropping the distinction between root and leaf, the group action extends to $\mathbb{S}_{n+1}$. On the other hand, the suspension of BHV_n^+ is $\mathbb{S}_n$-equivariantly homotopy-equivalent to a version [10, Section 9] of the geometric realization of the poset of partitions of n. The associated homology representations are fundamental in (among other things) the theory of operads.

3.2. A tree embedded in the plane inherits an orientation. To define coordinate patches on the space K_n of (isometry classes of) *planar* rooted metric trees with n leaves, labeled $\{1, 2, \ldots, n\}$ clockwise, we need to assign weights to the internal edges, and it is convenient to distinguish those weights from the lengths defined by a plane embedding; in particular, terminal edges are to be understood as having infinite weight.

This data defines a parametrization of the space such planar trees by a Catalan number

$$C_{n-1} = \frac{1}{n}\binom{2n-2}{n-1} = \frac{2^{n-1}}{n!}\,(2n-3)!! \tag{3.1}$$

of orthants $\mathbb{R}^{n-2}_{\geq 0}$, which glue together to define the *W-construction* of Boardman and Vogt [8]. Figure 3(a) shows the construction for $n = 4$ by glueing five quadrants (drawn as squares), along with a labeling of one of the squares in (b). In general, all the orthants will share a common vertex defined by the corolla.

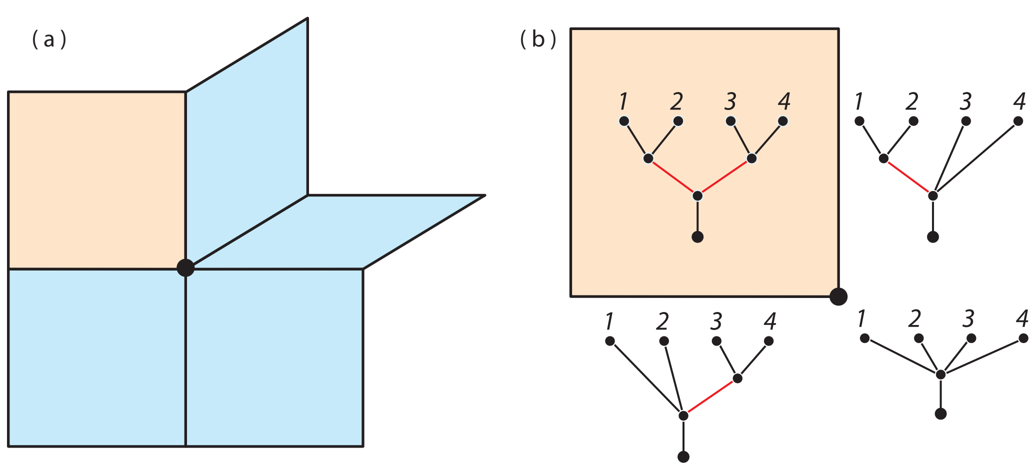

FIGURE 3. The piecewise-Euclidean geometry of K_4.

For any such tree there is a unique path ρ_k from the k-th leaf to its root. Let W_k be sum of the weights of the internal edges shared by the paths ρ_k and ρ_{k+1}. Define

$$v_k \;=\; \sum_{0<i<k} \exp(-W_i) \qquad \text{for } k \in \{2, \ldots, n\}\ ,$$

and let $v_1 = 0$; then the map

$$\mathfrak{d}\ :\ K_n \to \mathrm{Config}^n(\mathbb{R})/\mathrm{Aff}_+\ :\ T \mapsto \{0 = v_1 < v_2 < \cdots < v_n\ (< v_{n+1} = \infty)\} \tag{3.2}$$

embeds K_n in a space of configurations of points on the line.

Figure 4(a) shows an example with seven leaves having five (colored) internal edges. Here, we have $W_1 = d_1 + d_2$, $W_2 = d_2$, $W_3 = 0$, $W_4 = d_3 + d_4$, $W_5 = d_3$ and $W_6 = d_3 + d_5$. When the tree is a corolla (no internal edges) with n leaves, then $W_k = 0$ for all k. This corolla is mapped to the configuration $(0, 1, 2, \cdots, n-1)$ of points in $\mathbb{R}$. On the other

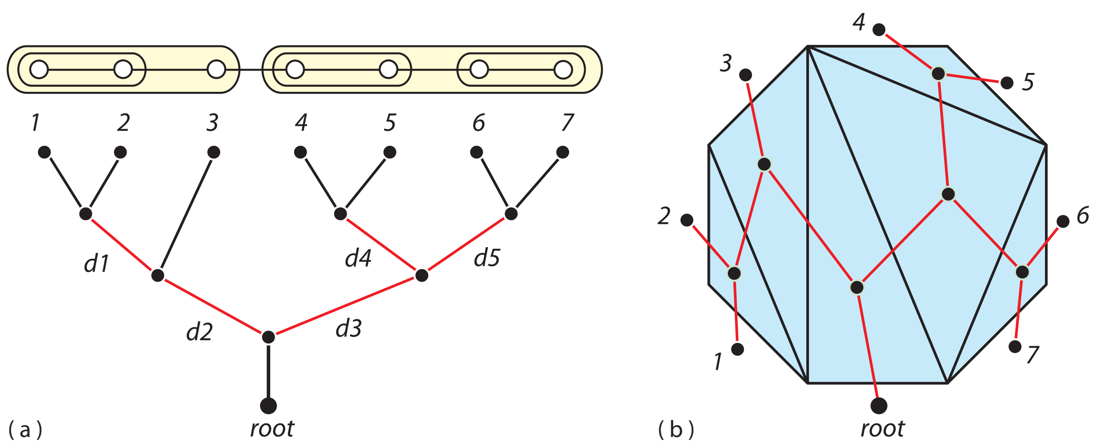

FIGURE 4. (a) Bracketings and planar trees and (b) polygons with diagonals.

hand, if $W_k \to \infty$, then $(v_{k+1} - v_k) \to 0$, and the corresponding points in the configuration space collide. Completing the orthants $\mathbb{R}^{n-2}_{\geq 0}$ to cubes $[0, \infty]^{n-2}$ defines the manifold-with-corners K_n of planar labeled trees, whose faces parametrize degenerate trees (with at least one weight equal to ∞). Their structure will be considered in greater detail below.

3.3. This space of planar rooted trees is Stasheff's polyhedron:

Theorem 3. *The pseudometric space K_n of planar labeled trees is homeomorphic to the* associahedron*: the convex polytope of dimension $n-2$ whose face poset is isomorphic to that of bracketings of n letters, ordered so $a \prec a'$ if a is obtained from a' by adding new brackets.*

These polytopes were constructed independently by Haiman (unpublished) and Lee [27], though Stasheff had defined the underlying abstract object twenty years previously, in his work on associativity in homotopy theory [34]. Figure 5(a) shows the 2D associahedron K_4 with a labeling of its faces, and (b) shows the 3D version K_5. There are over a hundred

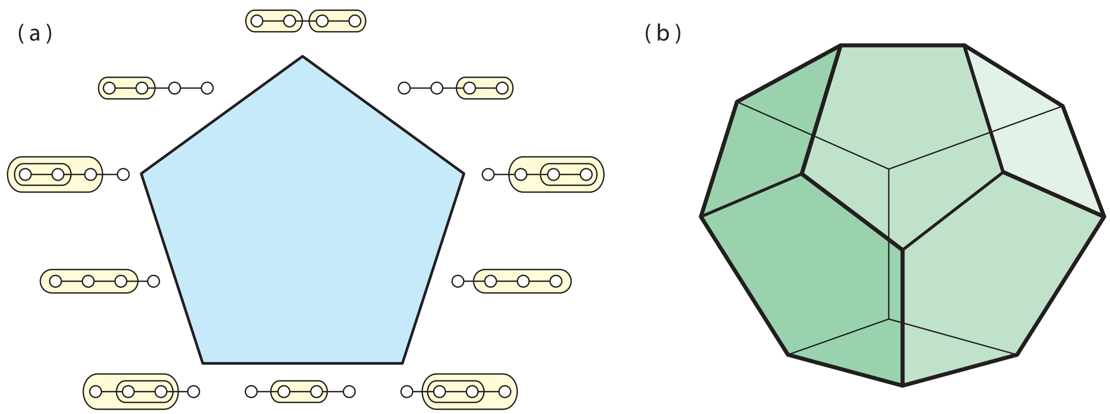

FIGURE 5. Associahedra K_4 and K_5.

combinatorial and geometric interpretations of the Catalan numbers, which index the vertices of the associahedra. Most important to us is the relationship between bracketings of n letters, which capture collisions of points on the line, and rooted planar binary trees. Figure 4(a) illustrates the bijection with rooted trees, while (b) shows the relationship to polygons with diagonals.

Remark. Considering trees as nested bracketings on the interval, as in Figure 4(a), allows a generalization of Equation (3.2), resulting in a map from *graph* associahedra to the minimal blow-up Coxeter complexes [9].

The Boardman-Vogt construction is a decomposition of the associahedron into cubes, exemplifying the general cubic barycentric subdivision of a polytope. Figure 6 shows the decomposition for the examples from Figure 5, where one cube from each complex has been highlighted. Each vertex of K_n, and hence each cube of the complex, is associated to a planar binary tree with n leaves; there is thus a Catalan number of cubes in the decomposition.

These spaces have some important symmetries. Relabeling the leaves of a tree in clockwise order (that is, turning over the piece of paper on which the tree is drawn) defines an involution $\varepsilon : K_n \to K_n$. A related construction interprets the root as a *zeroth* leaf, shifting the resulting indexing by a one-step clockwise rotation r. This defines an action of cyclic group $\mathbb{Z}_{n+1}$ on K_n which, together with the involutions above, extends to an

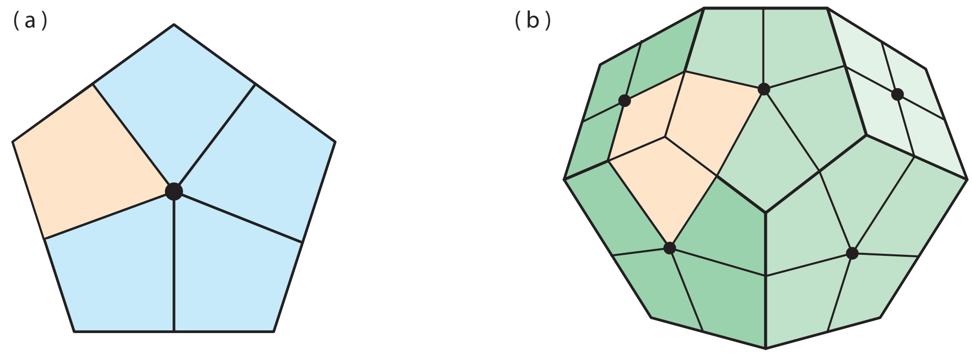

FIGURE 6. The cubical decomposition of associahedra.

action of the dihedral group D_{n+1}, presumably equivalent to that defined by Lee. This points to the usefulness of trees with arbitrarily labeled leaves, indexed by elements of $\mathbb{S}_n$ (if the root is distinguished) or by $\mathbb{S}_{n+1}$ (if not).

4. Compactified configuration spaces

4.1. The moduli problem for algebraic curves has been a central problem in mathematics since Riemann. It was solved in the 1970s by Deligne, Mumford, Knudsen [25] and others, over the integers $\mathbb{Z}$. A very special case of their general results constructs a moduli space for *real* algebraic curves of genus zero marked with $n > 3$ distinct smooth points. That solution can be regarded as a good compactification $\overline{\mathcal{M}}_{0,n}(\mathbb{R})$ of the space

$$\mathcal{M}_{0,n}(\mathbb{R}) \; = \; \mathrm{Config}^n(\mathbb{RP}^1)/\mathbb{P}\mathrm{Gl}_2(\mathbb{R})$$

of n distinct points on the real projective line. One of the objectives of this note is to provide this space with a natural metric, and thus an intrinsic measure.

Example. Figure 7(a) shows shows $\overline{\mathcal{M}}_{0,5}(\mathbb{R})$, obtained by blowing up three points on a 2-torus; part (b) shows $\overline{\mathcal{M}}_{0,6}(\mathbb{R})$ as a 3-torus, with the blowup along 3 points and 10 lines. For a detailed construction of these spaces as blowups of tori, see [13, Section 4].

4.2. The map $\mathfrak{d}$ of equation (3.2) embeds K_n in a configuration space of points on $\mathbb{R}$. Allowing arbitrary labelings of the leaves and root extends that construction to a surjection

$$\mathfrak{D} : K_n \times_{D_{n+1}} \mathbb{S}_{n+1} \; \to \; \overline{\mathcal{M}}_{0,n+1}(\mathbb{R})$$

to the moduli space for marked real genus zero curves, but this abbreviated description omits the details [13, Section 3.1] of the identifications among the faces of the associahedra. We reformulate that information here in the language of orbifolds.

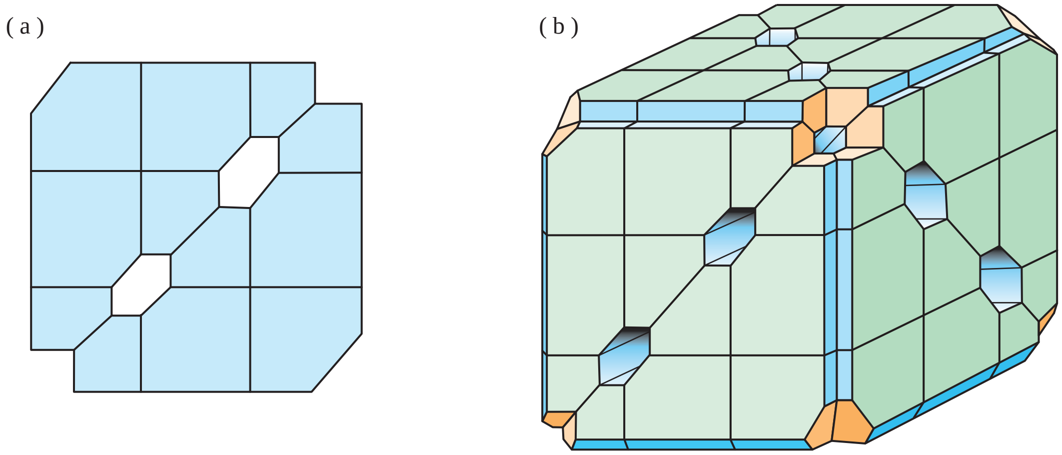

FIGURE 7. (a) $\overline{\mathcal{M}}_{0,5}(\mathbb{R})$ and (b) $\overline{\mathcal{M}}_{0,6}(\mathbb{R})$ as blowups of tori.

The polyhedral structure on K_n makes it a CW complex: a suitably topologized disjoint union of open cells corresponding to the interiors of its faces. The faces are indexed by degenerate (rooted, labeled) metric trees T (ie, with at least one internal edge of infinite weight).

Define the *degeneration* of T to be the disjoint union of subtrees T_e obtained by 'breaking' T along the edges e of infinite weight — each such edge then corresponds to a new leaf on each daughter tree. This defines an open cell

$$\text{Cell}_T \ := \ \prod_e \overset{o}{K}_{|T_e|}$$

for each face of K_n and hence a CW-decomposition

$$K_n \ = \ \coprod \text{Cell}_T \ .$$

With the morphisms defined below, the metric groupoid whose objects are pairs

$$T \times \sigma \ \in \ K_n \times \mathbb{S}_{n+1} \ .$$

will have a manifold as quotient space; but it will not an orbifold in the strictest sense (because its underlying space is a manifold with corners).

Remark. These planar trees can be viewed as *ribbon trees*, with a natural cyclic order around each vertex given by the planar embedding; see Figure 8(a). As the weight of an internal edge increases, we think of the thickness of the ribbon as decreasing, as in part (b). As the weight approaches ∞, the edge becomes a ribbonless segment.

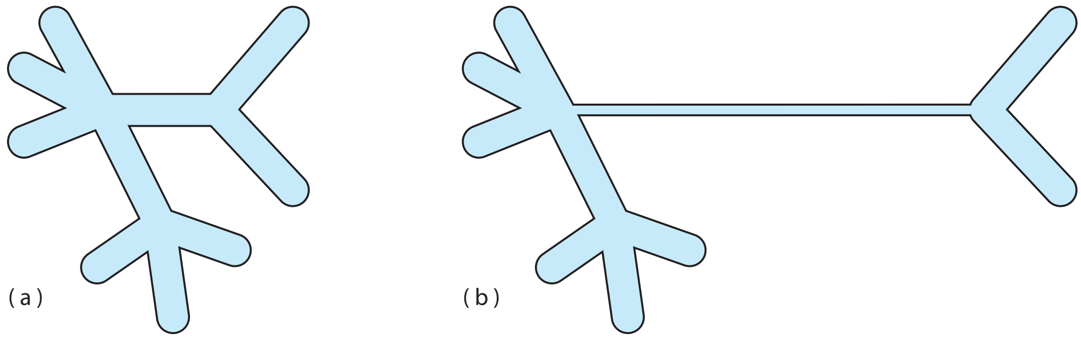

FIGURE 8. A ribbon tree as the weight of an internal edge is altered.

4.3. To define the morphisms of the category, note that any internal edge e of T partitions T into two subtrees, one containing the root. Let $e^*(T)$ denote the rooted planar tree obtained by reversing the embedding of the part of the tree away from the root. A planar rooted labeled tree is allowed to *twist* at an internal node v if the internal edge adjacent to v (closer to the root) has infinite weight. Figure 9 shows two examples of twists, where the green internal edge is degenerate.

The leaves of $e^*(T)$ inherit a natural order from its planar embedding, defining an element $\tau_e \in \mathbb{S}_{n+1}$ which sends the original labeling to the new one. This produces a homeomorphism

$$\mathrm{Cell}_T \times \sigma \ \rightarrow \ \mathrm{Cell}_{e^*(T)} \times (\sigma \circ \tau_e)$$

of the cell indexed by T with leaves labeled by σ, to the cell indexed by $e^*(T)$ with labeling $\sigma \circ \tau_e$. There is a unique smallest topological groupoid with objects as above, and these equivalences as (generating) morphisms. Its quotient manifold is the double cover of $\overline{\mathcal{M}}_{0,n+1}(\mathbb{R})$ defined by Kapranov [23]. If we adjoin the reflection around the root to the morphisms generating the category, we get $\overline{\mathcal{M}}_{0,n+1}(\mathbb{R})$ itself [14, Proposition 9].

There is an elegant construction of $\overline{\mathcal{M}}_{0,n+1}(\mathbb{R})$ and Kapranov's double cover based on blowups of $\mathbb{RP}^n$. The symmetric group $\mathbb{S}_n$ acts by reflections (ij) across the collection $\{x_i = x_j\}$ of hyperplanes in the subspace

$$x_1 + x_2 + \cdots + x_n = 0$$

of $\mathbb{R}^n$. This *braid arrangement* decomposes the $(n-2)$-sphere around the origin into $n!$ simplicial chambers, forming a type-A Coxeter complex. Kapranov showed that blowing up certain cells of this complex yields his double cover of $\overline{\mathcal{M}}_{0,n+1}(\mathbb{R})$; a combinatorial version of his construction is given in [14].

Example. Figure 10(a) shows the braid arrangement decomposing the 2-sphere into $4! = 24$ chambers. Part (b) shows the blowups along eight (nonnormal) crossings of

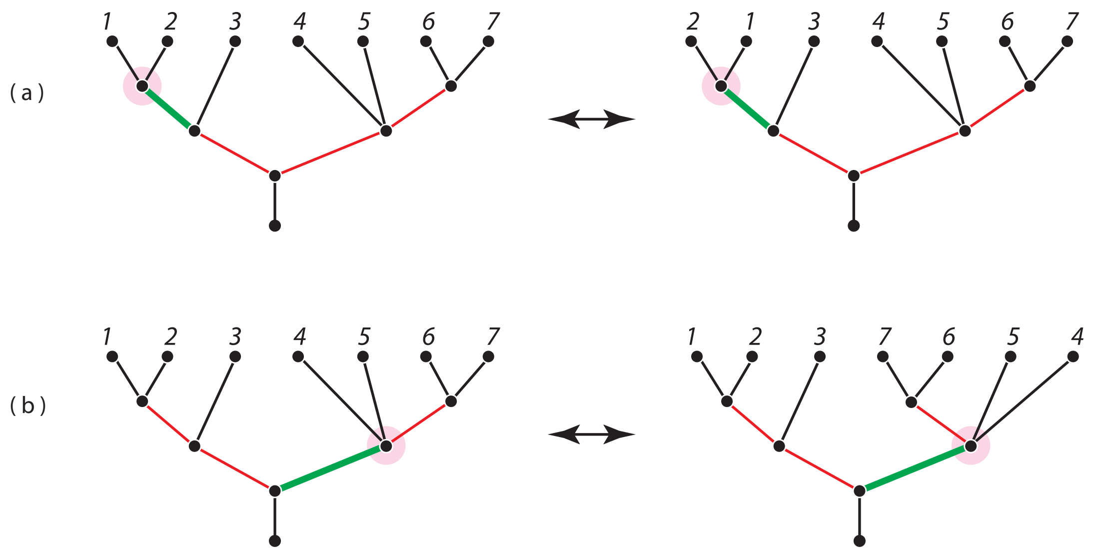

FIGURE 9. Two examples of twisting along degenerate edges.

this complex, where each hexagon is a crosscap, resulting in a nonorientable double cover of $\overline{\mathcal{M}}_{0,5}(\mathbb{R})$. Taking the antipodal map results in $\overline{\mathcal{M}}_{0,5}(\mathbb{R})$, shown in (c), tiled by 12 associahedra; compare this with Figure 7(a).

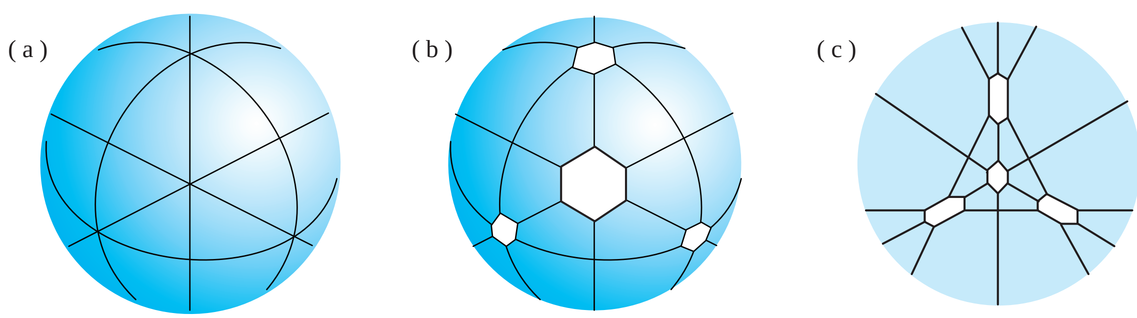

FIGURE 10. (a) Braid arrangement on the 2-sphere, (b) Kapranov's cover, and (c) the moduli space $\overline{\mathcal{M}}^{\text{or}}_{0,5}(\mathbb{R})$.

5. The orientable double covers

5.1. Since $\overline{\mathcal{M}}_{0,n+1}(\mathbb{R})$ is a nonorientable manifold if $n > 3$, it has an orientable double cover $\overline{\mathcal{M}}^{\text{or}}_{0,n+1}(\mathbb{R})$. In order to construct this double cover, we modify the morphisms in the discussion above. Let $\rho \in \mathbb{Z}_{n+1}$ denote a clockwise rotation

$$\mathcal{T}_{,} \to \rho^*(T)$$

of tree T by one step. Then, the category with the space $K_n \times \mathbb{S}_{n+1}$ of objects, and morphisms generated by twists along degenerate edges *and* the action ρ of $\mathbb{Z}_{n+1}$ defined by shifting the root (but *not* including twists around it) defines a presentation

$$K_n \times_{\mathbb{Z}_{n+1}} \mathbb{S}_{n+1} \to \overline{\mathcal{M}}^{\mathrm{or}}_{0,n+1}(\mathbb{R})$$

of the orientation cover of $\overline{\mathcal{M}}_{0,n+1}(\mathbb{R})$.

Example. The orientation cover $\overline{\mathcal{M}}^{\mathrm{or}}_{0,5}(\mathbb{R})$ can be constructed by regarding the hexagons in Figure 7(b) not as crosscaps but as "holes" through the sphere. The holes are then identified across the sphere, forming a surface of genus four. A construction of this cover with high symmetry is provided in [3], based on a pentagonal decomposition of Kepler's *great dodecahedron* shown in Figure 11.

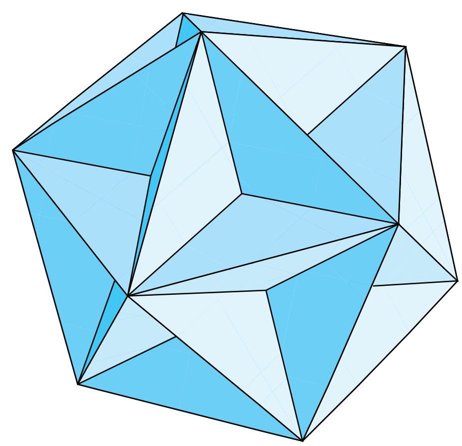

FIGURE 11. The great dodecahedron of Kepler.

5.2. The orientation cover $\overline{\mathcal{M}}^{\mathrm{or}}_{0,n+1}(\mathbb{R})$ is tiled by $n!$ associahedra, and hence, from equation (3.1), by $2^{n-1}(2n-3)!!$ cubes. Forgetting the extra data encoded by the planar structure o n its elements defines a smooth blowup

$$\psi \ : \ \overline{\mathcal{M}}^{\mathrm{or}}_{0,n+1}(\mathbb{R}) \ \to \ \mathrm{BHV}^+_n \ .$$

of the compactified Biller-Holmes-Vogtmann [6, Section 3.1] space by a compact smooth orientable aspherical manifold. Since $\overline{\mathcal{M}}^{\mathrm{or}}_{0,n+1}(\mathbb{R})$ is right-angled, 2^{n-2} associahedra meet at each vertex, and the center of each associahedron corresponds to a different labeling of the rooted planar corolla of n leaves. All such center vertices collapse into the unique vertex of BHV^+_n corresponding to the abstract rooted corolla. In fact, the map can locally be viewed as *origami flat folding* of an $(n-2)$-cube (cut into 2^{n-2} smaller cubes) into a cube in BHV^+_n, iteratively folding the cube in half and identifying the 2^{n-2} cubes as one.

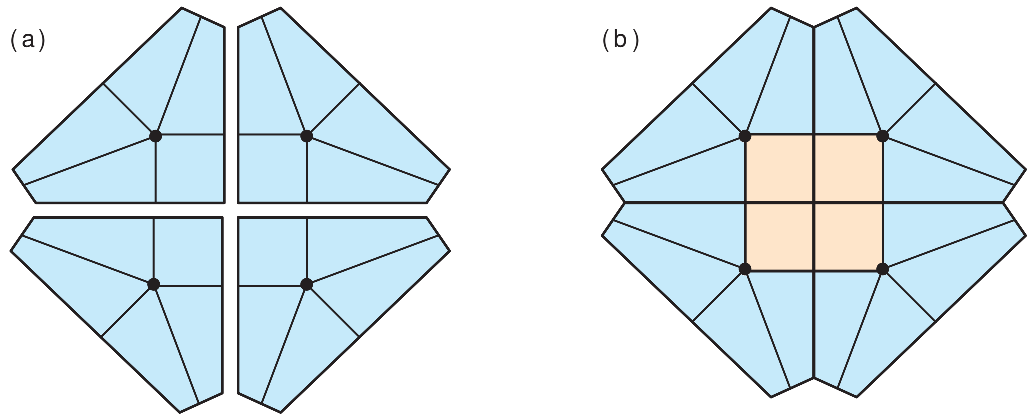

FIGURE 12. (a) Glueing associahedra in $\overline{\mathcal{M}}_{0,5}(\mathbb{R})$, with (b) cubes identified in BHV_4^+.

Example. Figure 12(a) shows four associahedra K_4 which will glue together, providing a local picture of $\overline{\mathcal{M}}_{0,5}(\mathbb{R})$. The cubical decomposition of each pentagon ensures that 2^2 smaller squares meet at a vertex of $\overline{\mathcal{M}}_{0,5}(\mathbb{R})$, as shown in the shading of Figure 12(b). The map can be visualized as folding of the dissected square (shaded region of Figure 12) into one square, as in Figure 13.

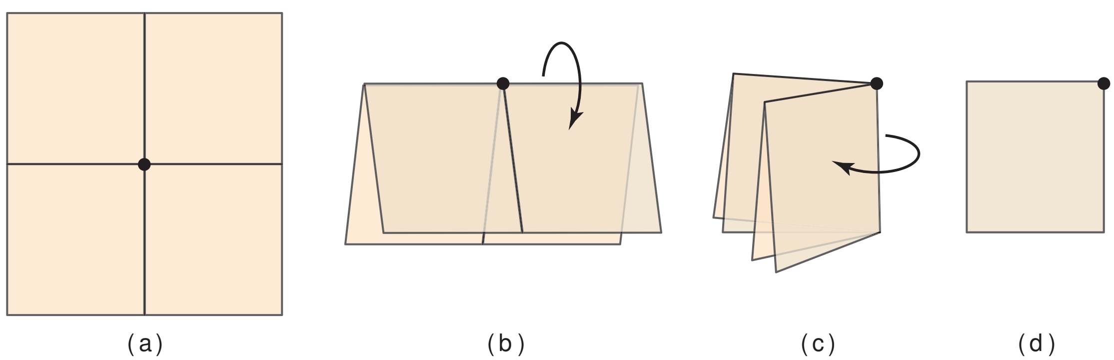

FIGURE 13. The $2^{n-2}:1$ folding of cubes.

Theorem 4. *The resulting $\mathbb{S}_{n-1}$-equivariant composition*

$$K_n \times_{\mathbb{Z}_{n+1}} \mathbb{S}_{n+1} \;\to\; \overline{\mathcal{M}}^{\mathrm{or}}_{0,n+1}(\mathbb{R}) \;\to\; \mathrm{BHV}_n^+ \;\simeq\; S^1 \wedge \mathcal{T}_n \;\sim\; S^{n-2} \wedge \mathbb{S}_{n-1}^+ \tag{5.1}$$

is 2^{n-2}-to-1 on the interiors of the cubes.

The space BHV_n is covered by $(2n-3)!!$ orthants, and the map forgets the order of the pairings in the Diaconis-Holmes labeling of binary trees. Dividing out the action of $\mathbb{S}_{n-1}$

in equation (5.1) defines a map

$$K_n \times (\mathbb{S}_n/\mathbb{S}_{n-1}) \;\to\; S^{n-2}\,;$$

this is just an $n = (\mathbb{S}_n/\mathbb{S}_{n-1})$-fold copy, parametrized by the cyclic relabelings of the leaves of trees in K_n, of the degree 2^{n-2} map

$$K_n/\partial K_n \;\to\; I^{n-2}/\partial I^{n-2} \;\simeq\; S^{n-2}$$

which collapses the boundary of the associahedron to a point, and folds the cubical decomposition as described above.

Example. The space $\overline{\mathcal{M}}_{0,4}(\mathbb{R})$ is a circle, obtained by gluing the three intervals $[0,1]$, $[1,\infty]$, $[\infty,0]$ end-to-end, and in this case our orientation cover is the double cover defined by gluing six copies of an interval in a circle. Our collapse maps this to a wedge $S^1 \wedge \mathcal{T}_3$ of three circles, which may be confusing since $\mathbb{S}_2^+$ consists of three points, but is the wedge $S^0 \vee S^0$ of *two* 0-spheres.

5.3. The piecewise Euclidean pseudometric defined by the W-construction blows up at degenerate trees, and its associated measure is not a probability measure, though it can easily be "mollified" (by some choice of cutoff) to be one. However, there may be something more canonical. For the two-dimensional case, the logarithmic derivative

$$\alpha(z_0 : z_1 : z_2 : z_3) := d\log[z_0 : z_1 : z_2 : z_3] = A_{01} - A_{02} + A_{23} - A_{13}$$

of the cross-ratio, where

$$A_{ik} \;:=\; (z_i - z_k)^{-1}(dz_i - dz_k) \;\in\; \Omega^1(\mathrm{Config}^n(\mathbb{C}))$$

is Arnol'd's one-form, defines a candidate

$$\omega(s,t) \;:=\; \alpha(1:0:\infty:s) \wedge \alpha(t:1:\infty:0)$$

for a volume form on $\overline{\mathcal{M}}^{\mathrm{or}}_{0,5}(\mathbb{R})$. Its integral over $\overset{o}{K}_4 = \{0 < s < t < 1 < \infty\}$ is the period

$$\int_{0<t<s<1} s^{-1}ds \;\wedge\; (1-t)^{-1}dt \;=\; \zeta(2) \;=\; -\frac{(2\pi i)^2}{24}$$

of a certain motive [2, Section 25.1] studied by Goncharov and Manin [19]. This is generalized by the map

$$\overline{\mathcal{M}}^{\mathrm{or}}_{0,n+1}(\mathbb{R}) \supset K_n \to \overline{\mathcal{M}}_{0,4}(\mathbb{R})^{n-2}$$

where

$$(0 < x_1 < \ldots x_{n-2} < 1 < \infty) \;\mapsto\; ([0:x_1:1:\infty],\ldots,[0:x_{n-2}:1:\infty])\;,$$

which pulls back the natural volume form on the target $(n-2)$-torus. The details of the symmetric group action are left to the reader.

6. Toward possible applications

6.1. The constructions above allow us to extend the diagram defining the space $\widetilde{\mathcal{Q}_n}$:

$$\begin{array}{ccc}
\widetilde{\mathcal{Q}_n} & \dashrightarrow & \mathcal{Q}_n \\
\downarrow & & \downarrow \\
\overline{\mathcal{M}}^{\mathrm{or}}_{0,n+1}(\mathbb{R}) & \longrightarrow & \mathrm{SP}^n(\mathbb{R})/\mathrm{Aff}_+ \sim \Delta^{n-2} \\
\downarrow & & \downarrow /\partial \\
\mathrm{BHV}^+_n \sim \mathbb{S}^+_n \wedge S^{n-2} & \xrightarrow{/\mathbb{S}_n} & S^{n-2}
\end{array}$$

(where $/\partial : \Delta^{n-2} \to S^{n-2}$ collapses the boundary of the simplex to a point). There are thus many possible variations on this theme, such as those involving blowups based on the Billera-Holmes-Vogtmann space as well as our resolution of it. We will return to this in part II.

Quite a lot is now known about the homotopy theory of the spaces $\overline{\mathcal{M}}_{0,n+1}(\mathbb{R})$: in particular, their tesselations imply their negative curvature in the sense of Alexandrov, and hence that they are $K(\pi, 1)$ manifolds, whose fundamental groups are the twisted right-angled Coxeter groups of the associahedra [12]. Generators and relations [22] are known for these groups, as is their cohomology [17, 31], which has some striking analogies with that of the pure braid groups. By a well-known principle in algebraic geometry [21], the mod two cohomology of $\overline{\mathcal{M}}_{0,n+1}(\mathbb{R})$ is isomorphic (after halving dimensions) to the mod two reduction of the cohomology of the corresponding complex algebraic varieties, while their rational cohomology is that of a product of planes with k^2 punctures, for $k \leq n/2 - 1$. Incidentally, the cohomology of the pure braid group on n strands is that of the product of copies of $\mathbb{C} - \{k\}$ for $k \leq n$.

6.2. One of the vexing problems in theoretical biology is the relation between genotype (easily measurable by DNA sequencing) and phenotype (less easily defined, or measured): this resembles old questions in quantum mechanics about observables and hidden variables. Although it is in many ways quite naive to say so, there may be interesting connections between models of evolution based on resonances in systems of linked oscillators, and the inverse problem of reconstructing evolutionary trees by dissimilarity matrix techniques. For a more general perspective, see [18]. In the language suggested here, this becomes a question about maps between moduli spaces of quadratic forms.

Biologists study rare events over enormous time-scales. Descent diagrams such as Darwin's Figure 1 go back to the beginning of modern thinking in the field: they represent incidence relations among experimentally-defined equivalence classes (species) in some hypothetical effectively infinite-dimensional stratified space of all viable organisms. Branching in descent diagrams can be modeled by specialization in the sense of algebraic geometry, defined (for example) by fixing some parameter. In the language of stratified spaces this corresponds to moving from the interior of some region to its boundary: something like a phase change (like water to ice). In such a cartoon description, a chicken is the Dirac limit of a tyrannosaur, as many of its genetic parameters are set equal to zero.

At this level of vagueness, there is reason to work with codimension than with probability: evolutionary events are highly unlikely, and in reasonable models will have effective probability zero; but in geometry any subspace of positive codimension has measure zero, and hence probability zero. The modern theory of phase change in condensed matter physics has developed powerful tools for the study of such transitions (viewed as moving to the boundary of some phenotype), but in current work there is usually only one such event in focus at a given moment. Evolution forces us to consider long concatenations of these events, and trees are a natural tool for their book-keeping. A geometric object with many strata (for example, a high-dimensional polyhedron, with faces of many dimensions) has an associated incidence graph, with a vertex for each face, and a directed edge between adjacent faces of lower dimension.[1] From this point of view, trees (and more generally graphs) can provide a kind of skeletal accounting of the relations between the components of geometric objects which are not as simply related as manifolds are to their boundaries.

Mathematicians are aware that objects of universal significance (such as symmetric groups) manifest themselves in unexpected contexts, and that their relevance to a subject can be signaled by the appearance of simpler but related objects (such as partitions, or Young diagrams). The immense utility of trees as a device for organizing evolutionary data points in this way toward configuration spaces in genomics; but questions there are so little understood that even very coarse models, such as those based on linear methods but with a few extra bells and whistles coming from astute compactifications, may provide useful insights. The dissimilarity matrices studied in phylogenetics have all entries nonnegative, with zeroes down the diagonal; they define a very interesting subspace of quadratic forms. The recent proof of a topological stability theorem [28, Remark 19] for the assignment of trees to dissimilarity matrices was one of the main motivations for this paper: it points toward the existence of a metric version, and fits very nicely in the framework presented here. We will return to this in part II.

[1] Similarly, mathematics monographs often include, as part of the introduction, a graph or *leitfaden* indicating the functional relations between subsections or chapters. Computer programs such as Mapper [29] provide something similar for data clouds.

References

1. P. Anderson and D. Stein. Broken symmetry, emergent properties, dissipative structures, life: are they related? in *Self-organizing systems*, Plenum Life Sciences Monograms (1987) 445–457.
2. Y. André. *Un introduction aux motifs*, Panoramas et Synthéses **17** (2004).
3. F. Apéry and M. Yoshida. Pentagonal structure of the configuration space of five points on the projective line, *Kyushu Journal of Mathematics* **52** (1998) 1–14.
4. V. I. Arnol'd. Modes and quasimodes, *Functional Analysis and its Applications* **6** (1972) 12–20.
5. S. Böcker and A. Dress. Recovering symbolically dated, rooted trees from symbolic ultrametrics, *Advances in Mathematics* **138** (1998) 105–125.
6. L. Billera, S. Holmes, K. Vogtmann. Geometry of the space of phylogenetic trees, *Advances in Applied Mathematics* **27** (2001) 733–767.
7. J. Boardman. Homotopy structures and the language of trees, in *Proceeding of Symposia in Pure Mathematics* **21** (1971) 37–58.
8. J. Boardman and R. Vogt. *Homotopy invariant algebraic structures on topological spaces*, Lecture Notes in Mathematics **347** (1973).
9. M. Carr and S. Devadoss. Coxeter complexes and graph-associahedra, *Topology and its Applications* **153** (2006) 2155–2168.
10. M. Ching. Bar constructions for topological operads and the Goodwillie derivatives of the identity, *Geometry and Topology* **9** (2005) 833–933.
11. M. Davis. Some aspherical manifolds, *Duke Mathematics Journal* **55** (1987) 105–139.
12. M. Davis, T. Januszkiewicz, R. Scott. Fundamental groups of blow-ups, *Advances in Mathematics* **177** (2003) 115–179.
13. S. Devadoss. Tessellations of moduli spaces and the mosaic operad, in *Homotopy Invariant Algebraic Structures*, Contemporary Mathematics **239** (1999) 91–114.
14. S. Devadoss. Combinatorial equivalence of real moduli spaces, *Notices of the American Mathematical Society* (2004) 620–628.
15. P. Diaconis and S. Holmes. Matchings and phylogenetic trees, *Proceedings of the National Academy of Sciences* **95** (1998) 14600–14602.
16. C. Epstein. Anderson localization, non-linearity and stable genetic diversity, *Journal of Statistical Physics* **124** (2006) 25–46.
17. P. Etingof, A. Henriques, J. Kamnitzer, E. Rains. The cohomology ring of the real locus of the moduli space of stable curves of genus 0 with marked points, *Annals of Mathematics* **171** (2010) 731–777.
18. M. Franceschet. PageRank: Standing on the shoulders of giants, available at `arXiv:1002.2858`
19. A. Goncharov and Y. Manin. Multiple ζ-motives and moduli spaces $\overline{\mathcal{M}}_{0,n}$, *Compositio Mathematica* **140** (2004) 1-14.
20. M. Gromov. Hyperbolic groups, in *Essays in group theory*, MSRI Publications **8** (1987) 75–263.
21. J. Hausmann, T. Holm, V. Puppe. Conjugation spaces, available at `arXiv:math/0412057`
22. A. Henriques and J. Kamnitzer. Crystals and coboundary categories, *Duke Mathematics Journal* **132** (2006) 191–216.

23. M. Kapranov. The permutoassociahedron, MacLane's coherence theorem, and asymptotic zones for the KZ equation, *Journal of Pure and Applied Algebra* **85** (1993) 119–142.
24. S. Keel and J. Tevelev. Equations for $\overline{\mathcal{M}}_{0,n}$, *International Journal of Mathematics* **20** (2009) 1–26.
25. F. Knudsen. The projectivity of the moduli space of stable curves II, *Mathematica Scandinavica* **52** (1983) 161–199.
26. H. Landau. The inverse eigenvalue problem for real symmetric Toeplitz matrices, *Journal of the American Mathematical Society* **7** (1994) 749–767.
27. C. Lee. The associahedron and triangulations of the n-gon, *European Journal of Combinatorics* **10** (1989) 551–560.
28. D. Levy and L. Pachter. The neighbor-net algorithm, *Advances in Applied Mathematics*, to appear; available at `arXiv:math/0702515`.
29. Mapper is available at `http://comptop.stanford.edu/programs/`
30. J. Morava. Braids, trees, and operads, available at `arXiv:math/0109086`.
31. E. Rains. The action of $\mathbb{S}_n$ on the cohomology of $\overline{\mathcal{M}}_{0,n}(\mathbb{R})$, *Selecta Mathematica* **15** (2009) 171–188.
32. A. Robinson and S. Whitehouse. The tree representation of $\mathbb{S}_{n+1}$, *Journal of Pure and Applied Algebra* **111** (1996) 245–253.
33. D. Speyer and B. Sturmfels. The tropical Grassmannian, *Advances in Geometry* **4** (2004) 389–411.
34. J. Stasheff. Homotopy associativity of H-spaces, *Transactions of the American Mathematical Society* **108** (1963) 275–292.

S. Devadoss: Williams College, Williamstown, MA 01267
E-mail address: `satyan.devadoss@williams.edu`

J. Morava: Johns Hopkins University, Baltimore, MD 21218
E-mail address: `jack@math.jhu.edu`