Searching combinatorial optimality using graph-based homology information

Pedro Real · Helena Molina-Abril · Aldo Gonzalez-Lorenzo · Alexandra Bac · Jean-Luc Mari

Abstract This paper analyses the topological information of a digital object O under a combined combinatorial-algebraic point of view. Working with a topology-preserving cellularization $K(O)$ of the object, algebraic and combinatorial tools are jointly used. The combinatorial entities used here are vector fields, V -paths and directed graphs. In the algebraic side, chain complexes with extra 2-nilpotent operators are considered. By mixing these two perspectives we are able to explore the problems of combinatorial and homological optimality. Combinatorial optimality is understood here as the problem for constructing a discrete gradient vector field (DGVF) in the sense of Discrete Morse Theory, such that it has the least possible number of critical cells. Fixing $\mathbb{Z}/2\mathbb{Z}$ as field of coefficients, by homological 'optimality' we mean the problem of constructing a 2-nilpotent codifferential map $\phi : C_*(K(O)) \to C_{*+1}(K(O))$ for finite linear combinations of cells in $K(O)$, called homology integral operator. The homology groups associated to the chain complex $(C(K(O)), \phi)$ are isomorphic to those of $(C(K(O)), \partial)$, being ∂ the canonical boundary or differential operator of the cell complex $K(O)$. Relations between these two problems are tackled here by using a type of discrete graphs associated to a homology integral operator, called Homological Spanning Forests (HSF for short). Informally, an HSF for a cell complex can be seen as a kind of combinatorial compressed representation of a homology integral operator. As main result, we refine the heuristic for computing DGVFs based on the iterative

The first two authors are partially supported by FEDER funds via the research projects FQM-296 from JJAA (Junta de Andalucía) and MTM2009-12716 from MICINN (Spain).

P. Real (✉) · H. Molina-Abril
Institute of Mathematics of University of Seville (IMUS), Seville, Spain
e-mail: real@us.es

A. Gonzalez-Lorenzo · A. Bac · J.-L. Mari
CNRS, LSIS UMR 7296, Aix-Marseille Université, Marseille, France

Morse complex reduction technique of [1], reducing the search space for an optimal DGVF to an HSF associated to a homology integral operator.

Keywords Discrete gradient vector field · Optimal discrete gradient vector field · Chain homotopy · Homological spanning forest · Homological information

1 Introduction

Most topological spaces of practical interest can be represented by a decomposition into subsets, each with simple topology, that are glued together along their boundaries. These decompositions are called cell complexes [2], and have been widely used for analysing topological spaces [3,4]. For instance, fixing 26-adjacency as the neighbouring connectivity for a digital object O, its connectivity can be encoded in terms of voxels (0-cells), edges between neighbouring voxels (1-cells), faces connecting three or four neighbouring voxels (2-cells) and volumes connecting up to eight mutually 26-adjacent voxels (3-cells).

The cell complex $K(O)$ of a given digital object O is susceptible to be analysed from a homological point of view.

Working with coefficients in $\mathbb{Z}/2\mathbb{Z}$, an algebraic topological representation related to a finite cell complex $K = (K_n)_{n\geq 0}$ is introduced in [1]. This representation, called *chain–integral complex* $\tilde{C}(K) = (\mathbb{Z}/2\mathbb{Z}[K_*], \partial, \phi)$, is described as follows:

(a) The elements of the graded vector space $\mathbb{Z}/2\mathbb{Z}[K_*]$ are finite linear combinations of cells of K;
(b) The linear map $\partial : \mathbb{Z}/2\mathbb{Z}[K_*] \to \mathbb{Z}/2\mathbb{Z}[K_{*-1}]$, called *differential operator*, is the canonical boundary operator of K (that is the linear map algebraically codifying the boundary of each set of cells);
(c) The linear map $\phi : \mathbb{Z}/2\mathbb{Z}[K_*] \to \mathbb{Z}/2\mathbb{Z}[K_{*+1}]$, called *homology integral operator*, is a (not unique) map specifying a strong relationship between the canonical chain complex $C(K) = (\mathbb{Z}/2\mathbb{Z}[K_*], \partial)$ associated to K and its homology.

The maps ∂ and ϕ are nilpotent of degree two, that is, $\partial\partial = 0$ and $\phi\phi = 0$. Moreover, if they satisfy $\partial\phi\partial = \partial$ and $\phi\partial\phi = \phi$, the homology of K with regards to the operator ∂ is isomorphic to the homology of K with regards to ϕ.

Classical homology algorithms (matrix reduction to Smith normal form [5], incremental algorithm of Delfinado–Edelsbrunner [6], DMT [7], Effective Homology [8], AT-model theory [9,10], etc.) can be used or adapted to obtain a homology integral operator ϕ. From now on, we refer to this problem of finding a homology integral operator for a finite cell complex as *homological optimality*. Discrete Morse Theory is a method in which the combinatorial nature of a homology integral operator is partially handled in terms of gradient paths and discrete vector fields. The fundamental result of this theory establishes that the cell complex $\mathcal{X}$ with an admissible (without circuits) discrete vector field is homotopy equivalent (in consequence, isomorphic on the level of homology) to a new complex $\mathcal{A}$ consisting of only the critical cells. The main problem here consists of finding an admissible discrete vector field as complete as possible, that is, covering the maximum number of cells of the complex. This is a

high computationally complex problem [11] that will be referred here as *combinatorial optimality*. On the other hand, the global combinatorial structure that underlies in some homology integral operators has recently been discovered under the umbrella of the Homological Spanning Forest (HSF) theory [1,12,13].

The HSF theory generalizes to higher dimensions the labelling connected components strategy using spanning trees. In fact, this kind of labelling process of high dimensional homology classes is performed using directed subgraphs (in general, they are not acyclic) on the connectivity graph of the cell complex. In [1], a homology-based heuristic for finding near optimal discrete gradient vector fields based on iterated Morse reductions is given.

In this paper, we propose a heuristics for combinatorial optimality based on the construction of a homology integral operator directly on the connectivity graph of the original cell complex.

The paper is organized as follows. In Sect. 2, the process of cellularization and an introduction on how to represent homological information (that related to homology generators and their relations between them) of digital objects using HSFs are explained. In Sect. 3, relations between combinatorial and homological optimality are presented. Section 4 focusses on the formal definition of HSFs. An algorithm for constructing HSFs is given in Sect. 5. Preliminary results on combinatorial optimality are drawn in Sect. 6. Section 7 is devoted to conclusions.

2 Cellularization and Homological information of digital objects

2.1 Cellularization

The cellularization technique presented in [14] is used here to obtain a cell complex (in fact, a polyhedral complex) of a given digital object. A formal definition of cell complexes can be seen in [15]. Given a cell complex embedded in $\mathbb{R}^n$, each q-dimensional cell σ has as geometric boundary $\partial(\sigma)$ that is a set of i-cells for $i < q$. The i-boundary of σ is the set of i-cells within its geometric boundary.

The connectivity graph of a cell complex K is the graph $G(K)$ where the nodes represent the cells of the complex, and the edges represent the incidence relation among the cells. We can also define the $(i, i+1)$-*connectivity graph* $(i = 0, 1, \ldots)$ $G^{(i,i+1)}(K)$, which is the subgraph of $G(K)$ induced by the nodes representing the i-cells and the $(i+1)$-cells. An example is shown in Fig. 1 where the boundary of the 2-cell A is the set $\{1, 2, 3, 4, a, b, c, d\}$. Its 0-boundary is $\{1, 2, 3, 4\}$ and its 1-boundary is $\{a, b, c, d\}$. If K is a convex cell complex (that is, K is formed by convex cells), a geometric realization of the graph $G(K)$ using as nodes the barycentres of the different cells can be comfortably established.

Given a digital object O represented by the set of black n-xels in a n-dimensional matrix with 0 and 1 values and a local adjacency relation between n-xels, the cellularization technique we use here is based on the idea that each element is a 0-cell of the cell complex $K^n(O)$. For $n = 2$ and considering the 4-adjacency for black pixels, this cellularization corresponds to a cubical cell complex $K_4^2(O)$ (see Fig. 2). The same occurs for $n = 3$ and considering the 6-adjacency between voxels.

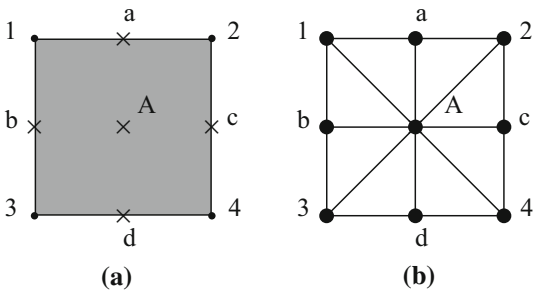


Fig. 1 **a** A cell complex K with four 0-cells (*1*, *2*, *3*, *4*), four 1-cells (*a*, *b*, *c*, *d*) and one 2-cell (A). **b** The connectivity graph $G(K)$

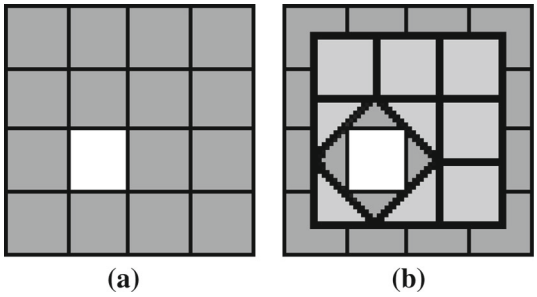


Fig. 2 **a** A two-dimensional digital object O and **b** its associated cell complex $K_8^2(O)$ (superimposed)

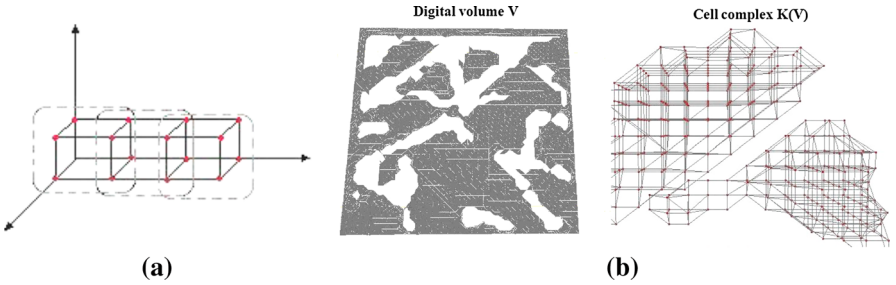


Fig. 3 **a** Points in *red* are voxels and the construction of $K_{26}^3(O)$ is gradually done for each $2 \times 2 \times 2$ *square* of voxels within the image. **b** Three dimensional digital object O and a zoom of a part of $K_{26}^3(O)$ (color figure online)

For $n = 3$ and considering the 26-adjacency for black voxels, 3-dimensional cells are formed by at most eight voxels mutually 26-adjacent. The dimension of a concrete cell will strongly depend on the geometric notion of point coplanarity in the euclidean space $\mathbb{R}^3$. For example, four voxels mutually 26-adjacent can form a tetrahedron or a two-dimensional square.

The cellularization process consists of dividing the object into overlapped entities. For $n = 3$, these entities are unit cubes formed by eight voxels mutually 26-adjacent (see Fig. 3). For $n = 2$, they are squares formed by four voxels. With overlapped, we mean that its intersection is a "square" of four voxels mutually 26-adjacent for $n = 3$, and an edge of two voxels for $n = 2$. The complex is scanned taking one entity per step, and obtaining its corresponding cell configuration. All the possible patterns for each voxel configuration within an entity have been previously computed and stored

into a look-up table (23 possible configurations within the unit cube, plus rotations). These patterns are computed in such a way, that when the configuration of each entity is glued together with its neighbouring entities to conform the cell complex of the entire object, intersections between simplexes are avoided.

2.2 Homological information

Given a cell complex, if we want to talk about homology, we first need to specify an "arithmetic", that is a field or a ring. In this paper, we work with coefficients in the field $\mathbb{Z}/2\mathbb{Z} = \{0, 1\}$. We need to bring to algebra the notion of cell complex K. This can be done by defining the canonical chain complex $(C(K), \partial_*)$ where ∂_i is the canonical i-boundary map $\partial_i : C_i(K) \to C_{i-1}(K)$, $i = 1, 2, \ldots$. This boundary map satisfies $\partial_i \partial_{i+1} = 0$, that is, the boundary of the boundary is zero.

We can now define the homology groups of a chain complex of the type $(C(K), d_*)$ (with differential map d_* not necessarily being the canonical one). An i-chain c (that is, a linear combination of i-cells) is a *cycle* if $d_i(c) = 0$. Any linear combination of 0-cells is, by definition, a 0-cycle, that is, $d_0(\sigma) = 0$ for every 0-cell σ. The q^{th} homology group of the chain complex $C(K)$ is $H_q(K) = \text{Ker } d_q / \text{Im } d_{q+1}, \forall q \geq 0$. In the three dimensional context, the Betti numbers are the ranks of the homology groups, and they are usually seen as the number of "different" holes of each dimension.

For example, the homology groups of the cell complex in Fig. 1 are all null excepting H_0, which is $\text{Ker } d_0 / \text{Im } d_1 = \mathbb{Z}/2\mathbb{Z}$. This result means that the complex has Betti numbers $\beta_0 = 1$ (one connected component) and $\beta_1 = 0$ (no holes).

Following the AT-model theory [9,10], we can derive homological information by determining a linear map (called homology integral operator) $\phi_* : C_*(K) \to C_{*+1}(K)$. The time complexity for computing a homology integral operator is $O(m^3)$ being m the number of cells. Now, from a canonical chain complex $(C(K), \partial_*)$ associated to a finite cell complex K, also endowed with an additional homology integral operator $\phi_* : C_*(K) \to C_{*+1}(K)$, it is easy to deduce homological information. In particular, the set of homologically different i-cycles is classified through the set of critical chains of the kind $(\mathbb{1} + \partial_{i+1}\phi_i + \phi_{i-1}\partial_i)(\sigma) = \sigma + (\partial_{i+1}\phi_i + \phi_{i-1}\partial_i)(\sigma)$, $\forall$ i-cell σ, $\forall i = 0, \ldots, n$.

In general, a homology integral operator is, computationally speaking, an algebraic artefact that holds too much redundancy. First steps towards the simplification of this operator are taken in [1], where HSFs are introduced. More precisely, HSF is a special set of direct graphs on the connectivity graph $G(K)$, that generalizes the notion of V-path in DMT [7]. Obtaining a homology integral operator starting from an HSF is, in general, an easy task. The reciprocal result of getting an HSF structure (non necessarily on $G(K)$) from a homology integral operator seems to be impossible in some occasions. The concept of HSF will be introduced in Sect. 4. Given a cell complex, an algorithm for constructing an HSF is presented in Sect. 5.

An example of HSF can be seen in Fig. 4 where the directed edges of the 0–1 HSF are coloured in red (from 0-cells to 1-cells) and orange (from 1-cells to 0-cells). The (1–2)-HSF is drawn in yellow. There is only one directed edge from the 1-cell $\{a, c\}$ to the 2-cell. The associated integral homology operator $\phi_* : C_*(K) \to C_{*+1}(K)$ is

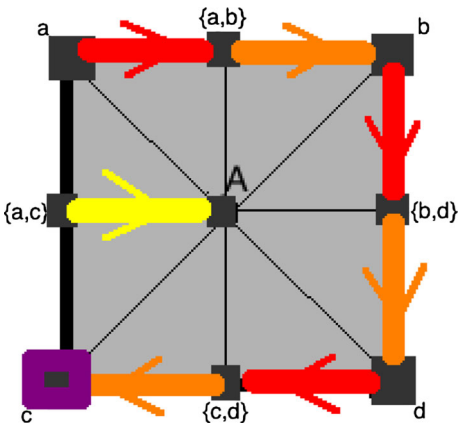


Fig. 4 An HSF structure for the two-dimensional cell complex in Fig. 1

defined by $\phi_0(a) = \{a, b\} + \{b, d\} + \{c, d\}$, $\phi_0(b) = \{b, d\} + \{c, d\}$, $\phi_0(d) = \{c, d\}$, $\phi_0(c) = 0$, $\phi_1(\{a, c\}) = A$, $\phi_1(\sigma) = 0$, for $\sigma \neq \{a, c\}$ and $\phi_2(A) = 0$. We can see that this operator satisfies the three conditions of the homology integral operator, and the set of representative cycles of homology generators reduces to the cell c for H_0.

3 Combinatorial and homological optimality

There exist a wide variety of perspectives from which homological information has been treated in the literature. Besides the classical view of homology as an algebraic-topological invariant, there are two other views that are of interest to this work:

(a) Homology as a representational tool for cell complexes (see, for example, [8–10,13]).
(b) Homology can be seen as a homotopical reduction of a cell complex to its minimal homological expression in terms of critical cells [16].

In DMT [7], the main homotopy operations in this reduction process are *simplicial collapses*. A simplicial collapse is graphically represented as a vector between an i-cell and an $(i + 1)$-cell. Combinatorial notions like V-paths and *discrete gradient vector fields* are used as an attempt to control the homology–homotopy reduction process. Having a complete control of this process is not possible without taking more complex homotopy operations into consideration. Results of the kind "contractibility does not imply collapsibility" [17] support this thesis. These complex homotopy operations are geometric realizations of elementary algebraic operations appearing in the construction of a homology integral operator and they can be expressed in graphical terms by elementary directed graphs. Some preliminary efforts to clarify this aspect have been the introduction of the chain–integral complex algebraic notion in [1] and the Homological Spanning Forest concept in [12,13].

Within the context of DMT, having combinatorial optimality for a finite cell complex K means that a discrete gradient vector field for K having the least number of critical cells in each dimension can be constructed. That means to have an acyclic matching between pairs of incident cells, where the number of unpaired cells is the least possible.

By homological optimality we mean here to construct an algebraic homology integral operator for the canonical chain complex $(C_*(K), \partial_*)$ in the sense of the AT-model theory.

Finding combinatorial optimality is a complex problem [11,18] and, in general, the suitable question to be solved consists of determining the degree of combinatorial near-optimality of K. On the other hand, homological optimality can be solved in general in $O(m^3)$ (being m the number of cells in K) but in most of the cases dealing with this problem, geometrical and combinatorial tools are not jointly used.

The idea presented here consists of expressing homological optimality using exclusively directed subgraphs on the connectivity graph $G(K)$ of the cell complex K. In this way, we obtain a directed subgraph of $G(K)$, having exactly β_i (corresponding with the Betti numbers) critical nodes (with no arrow coming out from them) in each dimension. This scaffolding will be our searching space for finding combinatorial near-optimality.

In [1], a recursive technique for constructing nested Morse complexes using directed trees and chain–integral complexes is designed. A homology integral operator ϕ is finally derived from that technique. Although the initial directed tree is forced to follow directed graphs on the connectivity graph, the final graph "describing" a homology integral operator obtained in [1] is not in general a subgraph of $G(K)$ and, consequently, it is not appropriate for efficiently searching combinatorial near-optimality.

Our goal in this paper is to design an algorithm that corrects this flaw by using graph-search algorithms on $G(K)$.

4 Defining optimal homological scaffoldings

We present an algorithm whose input is a finite cell complex K, and outputs a homology integral operator $\phi_* : C_*(K) \to C_{*+1}(K)$. As we have already commented, this integral operator can be deduced from a hierarchical set of directed graphs on the connectivity graph $G(K)$, called HSF of K. Let us suppose that the dimension of K is $|K| = k$ and that K is a cell complex such that any two $(i+1)$-cells share only one i-cell $\forall i \geq 0$. This kind of cell complex has been exhaustively used for describing digital images in topological terms.

Let us give a review of the main notions related to directed graphs. A directed graph $G = (V, E)$ is defined by a set of nodes and a set of (ordered) pairs of nodes (called directed edges, arcs or arrows). If (u, v) is an arc, u is the *tail* and v is the *head*. A direct path within the connectivity graph $G(K)$ from one cell σ to another τ is denoted by $\sigma \to \tau$. Let be $\mathcal{G}$ a directed subgraph on the connectivity graph $G(K)$ and σ one of its nodes, we call the directed subgraph generated by σ (and we denote it by $\mathcal{G}(\sigma)$) the subgraph induced by the nodes τ which have a (directed) path $\sigma \to \tau$ from σ to τ. In particular, if σ is an i-cell node and τ is an $(i+1)$-cell node, then the directed path $\sigma \to \tau$ runs in an alternating manner over i-cell nodes and $(i+1)$-cell nodes.

Now, we are able to define a MSF (Morse homological spanning forest) structure.

Definition 1 A MSF $\mathcal{M}$ for a finite cell complex K satisfies the following conditions:

(a) $\mathcal{M}$ is a directed graph.

(b) As non-directed graph, $\mathcal{M}$ is a subgraph of the connectivity graph $G(K)$.
(c) Any (weakly) connected component of $\mathcal{M}$ has as nodes cells of dimension i and $(i+1)$ and it is called an $(i, i+1)$-*MSF subgraph* of $\mathcal{M}$, for $i = 0, 1, \ldots, k$. Exceptionally, this subgraph can be formed by only one i-cell. The directed edges (σ, τ) from an i-cell σ to an $(i+1)$-cell τ in $\mathcal{M}$ are called *integral arrows*. The rest of the arrows from an $(i+1)$-cell to an i-cell are called *differential arrows*.
(d) Any cell of K must belong exactly to one $(i, i+1)$-MSF subgraph of $\mathcal{M}$.

Let us note the asymmetry in the definition of the following two algebraic maps associated to a MSF structure.

Definition 2 Given a MSF $\mathcal{M}$ for a finite cell complex K, the *associated differential operator* $d(\mathcal{M})_* : C_*(K) \to C_{*-1}(K)$ (resp. the *associated combinatorial integral operator* $\psi(\mathcal{M})_* : C_*(K) \to C_{*+1}(K)$) is a linear map decreasing the dimension by one (resp. increasing the dimension by one). For any $(i, i+1)$-MSF subgraph $\mathcal{J}$ of $\mathcal{M}$, let $\{\sigma_1, \sigma_2, \ldots, \sigma_r\}$ be the set of i-cell nodes of $\mathcal{J}$ and let $\{\tau_1, \ldots, \tau_s\}$ be the set of its $i+1$-cell nodes, then:

$$d(\mathcal{M})(\sigma_j) = 0; \quad \left(\text{resp. } \psi(\mathcal{M})(\sigma_j) = \sum_{i=1}^{s} b_{j,i}\tau_i\right)$$

$$d(\mathcal{M})(\tau_j) = \sum_{i=1}^{r} a_{j,i}\sigma_i; \quad (\text{resp. } \psi(\mathcal{M})(\tau_j) = 0)$$

where $a_{j,i}$ is 1 if (σ_i, τ_j) is an integral arrow or (τ_j, σ_i) is a differential arrow (resp. $b_{j,i}$ is 1 if (σ_j, τ_i) is an integral arrow) of $\mathcal{M}$ and 0 elsewhere.

An algebraic operator on $C_*(K)$ can be associated to an acyclic MSF $\mathcal{M}$ (in which all the $(i, i+1)$-MSF subgraphs are directed graphs without directed cycles).

Definition 3 Given an acyclic MSF $\mathcal{M}$ for a finite cell complex K, the *associated HSF-integral operator* $\phi(\mathcal{M})_* : C_*(K) \to C_{*+1}(K)$ is a linear map increasing the dimension by one and defined as follows. For any $(i, i+1)$-MSF subgraph $\mathcal{H}$ of $\mathcal{M}$, let $\{\sigma_1, \sigma_2, \ldots, \sigma_r\}$ be its i-cell nodes and let $\{\tau_1, \ldots, \tau_s\}$ be its $(i+1)$-cell nodes. Then:

$$\phi(\mathcal{M})(\sigma_j) = \sum_{k} b_{j,k}\gamma_k$$

$$\phi(\mathcal{M})(\tau_j) = 0$$

where $b_{j,k}$ is the sum of the number of directed paths in $G(K)$ from σ_j to any $(i+1)$-cell γ_k of K, having all the arrows from $\mathcal{M}$ possibly excepting some differential arrows that have not γ_k as tail.

It is immediate to prove that the integral operator $\phi(\mathcal{M})$ associated to a MSF $\mathcal{M}$ is 2-nilpotent, that is, it satisfies $\phi(\mathcal{M})_{*+1}\phi(\mathcal{M})_* = 0$. On the other hand, the graded set $Cr(\mathcal{M})$ of critical nodes of $\mathcal{M}$ is formed by those cells which are not the tail of an integral arrow of $\mathcal{M}$. A critical cell c satisfy that $\phi(\mathcal{M})(c) = 0$. In order to promote

an acyclic MSF $\mathcal{M}$ for K to an HSF directed graph for K, some additional algebraic conditions relative to its associated integral operators must be satisfied.

Definition 4 ([1]) An *HSF directed graph* $\mathcal{H}$ for a finite cell complex K of dimension k is an acyclic MSF structure such that its associated integral operator $\phi(\mathcal{H})$ satisfies the following conditions:

1. SDR-condition of $\mathcal{M}$ with regards to the boundary differential ∂: $\phi(\mathcal{H})_i \partial_{i+1} \phi(\mathcal{H})_i (\sigma) = \phi(\mathcal{H})_i(\sigma), \forall\, i\text{-cell}\ \sigma, \forall i = 0, 1, \ldots, k-1$;
2. Homology condition of $\mathcal{M}$ with regard to ∂: $\partial_i \phi(\mathcal{H})_{i-1} \partial_i(\sigma) = \partial_i(\sigma), \forall\, i\text{-cell}\ \sigma, \forall i = 1, \ldots, k$.

The HSF acronyms, coming from Homological Spanning Forest, is due to the fact that these directed graphs (in general, they are not forests) generalize to higher dimension the role of the spanning forest notion for labelling connected components. Informally speaking, an HSF directed graph structure is a topological and geometrical (coordinated-based) gadget which is mainly used here as a tool that facilitates the computation of the associated integral operator satisfying the three previous mentioned conditions. In fact, an HSF can be seen as a kind of compressed combinatorial coding of its associated integral operator. In principle, this notion heavily depends on the ring of coefficients we work with.

There are two kind of critical cells for a given HSF $\mathcal{H}$. Given a $(i, i+1)$ MSF subgraph of $\mathcal{H}$, a critical cell τ of dimension $(i+1)$ is called (homologically) *inessential* and a critical cell σ of dimension i is called *essential*. If $\mathcal{H}$ is simply a MSF structure, then τ is the head of an integral arrow and σ is the head of all the edges incident to it (differential arrows).

One algebraic integral operator can have associated two different HSFs. For instance, if we add to an HSF structure $\mathcal{H}$ all the differential arrows of the kind (τ, σ), being σ any essential critical i-cell of a $(i, i+1)$ MSF subgraph of $\mathcal{H}$ and τ a $(i+1)$-cell belonging to the same subgraph, then the resulting MSF structure is again an HSF, denoted by $\widetilde{\mathcal{H}}$, endowed with the same integral operator $\phi(\mathcal{H})$.

Let us note than when we talk about MSFs or HSFs of a digital object O, we mean MSFs or HSFs of its associated cellularization $K(O)$. Given an HSF $\mathcal{H}$ with its associated homology integral operator $\phi(\mathcal{H})$, the cycles given by the formula $\sigma + \partial\phi(\mathcal{H})(\sigma) + \phi(\mathcal{H})\partial(\sigma)$ (where σ runs over all the cells of K), are representative cycles of the homology generators of K. The same set of representative cycles can be calculated if we restrict σ to the smaller set $Cr(\mathcal{M})$.

In order to compute numerical or algebraic information related to homology generators and their relations between them at both homology and cycle levels, the importance of the notion of homology integral operator is great. In order to use this tool for progressing in computing advanced topological invariants, a successful homological algebra framework provided by integral–chain complexes is proposed in [12].

4.1 More examples of HSF

We give here some simple examples of the previously introduced concepts.

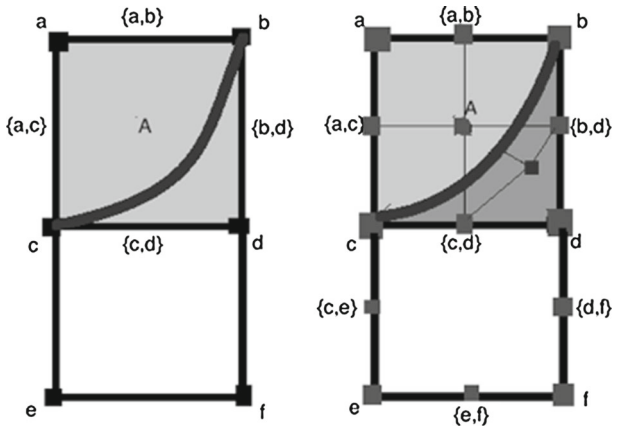


Fig. 5 A cell complex and a subgraph of its connectivity graph

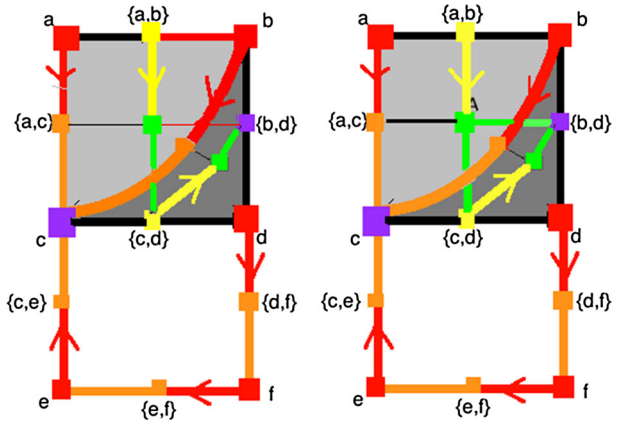


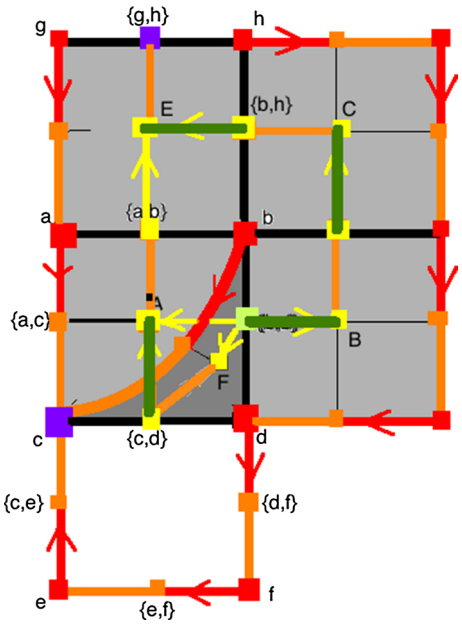
Fig. 6 A MSF graph and an HSF

In Fig. 5 (left) a cell complex K formed by two 2-cells (a square $abcd$ and a triangle bcd), eight 1-cells and six 0-cells is presented. On the right side, superimposed on K, we show a subgraph of $G(K)$ in which only edges connecting cells which differ in one dimension are drawn.

In Fig. 6 (left) we show a MSF directed graph $\mathcal{H}$ which is an HSF for the cell complex K of Fig. 5. There is one (0–1)-MSF subgraph having only one critical cell (the 0-cell c) and one (1–2)-MSF subgraph having only one critical cell (the 1-cell $\{b, d\}$). Let us note that the gradient vector field specified by the integral arrows of this HSF is already optimal.

In Fig. 6 (right) the HSF-structure $\tilde{\mathcal{H}}$ is drawn. The only difference with regards to the previous MSF is the green arrow from the square 2-cell A to the edge $\{b, d\}$. The number of critical cells is also two. Using the homology integral operator associated to this HSF we can determine the representative path $c = \{b, c, e, f, d, b\}$ of the homology generator associated to the critical cell $\{b, d\}$. In fact, c is the 1-cycle

Fig. 7 A non-optimal DGVF



$1 + \phi\partial(\{b, d\})$, where ϕ is the homology integral operator associated to the previous HSF. It can be noted that the number of optimal DGVFs that can be determined from $\mathcal{H}$ is three and from $\tilde{\mathcal{H}}$ is four.

Let us consider now the example shown in Fig. 7 where an HSF is shown. We can now try to obtain combinatorial optimality from the previous HSF by taking a maximal subset of pairwise disjoints arrows. A non-optimal DGVF is given by:

(a) the red integral arrows at level 0–1;
(b) at level 1–2 the set of yellow integral arrows, excluding $(\{b, d\}, A)$, $(\{b, d\}, F)$ and $(\{a, b\}, E)$.

This last set is shown in Fig. 7 where the level 1–2 set is coloured in green.

Combinatorial optimality is obtained by choosing the DGVF within the HSF (in green) in Fig. 8.

5 Constructing HSFs from MSFs

The cell complexes (K, ∂) we consider here are cellularizations of nD digital objects. Let us start determining a concrete kind of MSF, from which we progress for getting an HSF structure.

Definition 5 A non-directed tree $\mathcal{T}$ on the connectivity graph $G(K)$ is called a *collapsible* $(i, i+1)$ *tree* (*c-tree* for short) if it exclusively involves i-cells and $(i+1)$-cells and satisfies that every leaf of $\mathcal{T}$ is an i-cell. A *maximal non-directed* $(i, i+1)$ *c-tree* $\mathcal{T}$ is a $(i, i+1)$ c-tree containing the maximal number of $(i+1)$-cell nodes such that all their boundary i-cells nodes belong to $\mathcal{T}$.

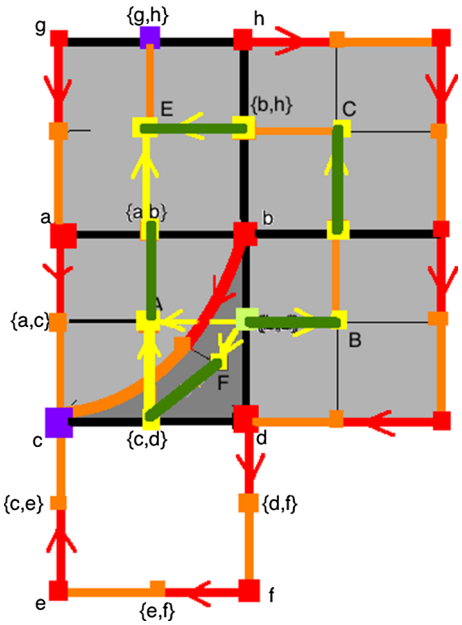


Fig. 8 An optimal GVF

Let us recall that the number of $(i+1)$-cells in the c-tree $\mathcal{T}$ sharing an i-cell in $\mathcal{T}$ with any other $(i+1)$-cell node is at most two. In other words, the valence of a i-cell node is at most two.

A $(i, i+1)$ c-tree $\mathcal{T}$ can be converted into a (non-unique) special directed acyclic graph (DAG, for short) with the minimum number (at least one) of essential critical i-cells and such that its associated integral operator $\phi(\mathcal{T})$ satisfies both the SDR and homology conditions with regards to its associated differential operator $d(\mathcal{T})$. This is called a *Morse directed* $(i, i+1)$ *c-tree associated to* $\mathcal{T}$. An constructive way to create such a directed tree is the following. We first set a root in $\mathcal{T}$ by selecting an i-cell μ. For any $(i+1)$-cell τ of the $2k$-level ($k = 1, \ldots$), we choose one incident i-cell σ of the $(2k+1)$-level and then stablish the integral arrow (σ, τ). Finally, differential arrows are set for the rest of the edges of $\mathcal{T}$. In DMT terms, the constructed DAG, denoted $\widehat{\mathcal{T}}$, is collapsible to a set of i-cells including μ.

Definition 6 A MSF $\mathcal{M}$ for a finite cell complex K of dimension k such that any (weakly) connected $(i, i+1)$-MSF subgraph is a Morse directed $(i, i+1)$ c-tree (resp. a Morse directed maximal $(i, i+1)$ c-tree) is called a *c-MSF* (resp., a *m-MSF*).

We construct an example of c-MSF structure in the following way. If any directed $(i, i+1)$ tree $\mathcal{D}$ of a MSF $\mathcal{M}$ is of the form $\mathcal{D} = \widehat{\mathcal{T}}$, being $\mathcal{T}$ a non-directed $(i, i+1)$ c-tree (resp. maximal non-directed c-tree), then $\mathcal{M}$ is a c-MSF (resp. a m-MSF).

In any directed $(i, i+1)$ tree $\mathcal{D}$ of a c-MSF $\mathcal{M}$, every $(i+1)$-cell node is the head of one integral arrow and every i-cell is the tail of at most one integral arrow of $\mathcal{M}$.

Let us note that the construction of m-MSF strongly depends on the particular order we choose for listing the set of cells of the complex. Any m-MSF $\mathcal{M}$ of K is an hierarchical DAG which at (0, 1) level is generated by a spanning forest of its 0-cells.

All the edges in the connectivity graph $G(K)$ which do not belong to a m-MSF $\mathcal{M}$ are edges either connecting a $(i, i+1)$ m-MSF tree with itself, with a $(i-1, i)$ m-MSF tree or a $(i+1, i+2)$ m-MSF tree ($i = 1, \ldots, k-2$). The associated differential operator $d(\mathcal{M}) : C_*(K) \to C_{*-1}(K)$ (see Definition 2) is zero for all the i-cells of a $(i, i+1)$ m-MSF tree. The associated integral operator $\phi(\mathcal{M})_* : C_*(K) \to C_{*+1}(K)$ (see Definition 3) is zero for all the $(i+1)$-cells of a $(i, i+1)$ m-MSF tree. It satisfies the SDR condition both with regards to $d(\mathcal{M})$ (see Definition 2) and to δ, the canonical boundary map ∂ of K restricted to the set of $(i+1)$-cells (delta is supposed to be zero over the i-cells) of the directed $(i, i+1)$ m-MSF trees of $\mathcal{M}$. It also satisfies the homology condition with regards to $d(\mathcal{M})$.

Finally, starting from a m-MSF $\mathcal{M}$ for the cell complex K, it is possible to define another MSF structure with additional homology-related properties using the same underlying set of c-trees of $\mathcal{M}$.

Definition 7 A m-MSF DAG structure whose associated integral operator $\phi(\mathcal{M})$ satisfies the homology condition with regards to the differential operator $\delta : C_*(K) \to C_{*-1}(K)$ is called *homotopy m-MSF* or *h-MSF* for short.

Let us show now a method for constructing an h-MSF from a m-MSF $\mathcal{M}$. From any Morse directed $(i, i+1)$ c-tree $\mathcal{M}_{(i,i+1)}$ of $\mathcal{M}$ having r ($r \geq 1$) essential critical i-cells, we construct another Morse directed c-tree in the following way. For each $(i+1)$-cell τ of $\mathcal{M}_{(i,i+1)}$ and for each i-cell σ of the combinatorial boundary of τ such that the edge $\{\sigma, \tau\}$ does not belong to $\mathcal{M}_{(i,i+1)}$, the idea is simply that σ becomes an essential critical cell in the current step of the loop. Taking into account that we handle with a c-tree, there always must be at least one directed path $\sigma \to \gamma$, being γ an essential critical i-cell in the current stage. A new path $\gamma \to \sigma$ reversing the arrows (differential arrows become integral and viceversa) of the previous path is constructed and σ turns into a critical cell in the resulting Morse directed $(i, i+1)$ c-tree at this step. At the end of this algorithmic process we always achieve a h-MSF, no matter the order in which we process the set of cells in the nested loop.

Let us consider a finite cell complex K of dimension k. There are several graph-based strategies that can be used to construct spanning trees for finding connected components: depth-first search, breadth-first-search, greedy algorithms, etc. We can use them using any order for the set of cells of K in order to determine the underlying hierarchical set of non-directed trees of a h-MSF over $G(K)$.

We are interested here in designing an algorithm for computing an HSF structure over a finite cell complex K starting from a h-MSF structure. We use one of the previous specialized spanning tree techniques for achieving the preprocessing step of constructing a h-MSF $\mathcal{M} = \{\mathcal{M}_k^{(i,i+1)}\}_{i,k}$. Due to the fact that $\mathcal{M}$ is a h-MSF, let us recall that the associated integral operator $\phi(\mathcal{M})$ satisfies the 2-nilpotency condition and the SDR and homology conditions with regards to $d(\mathcal{M})$ and the canonical boundary operator ∂ restricted to the set of $(i+1)$-cells (applied over an i-cell is zero) of each $(i, i+1)$ h-MSF tree of $\mathcal{M}$.

The main idea consists of "perturbing" the differential operator $d(\mathcal{M})$ in order to reach a new chain complex $(C_*(K), \partial))$ endowed with an HSF structure.

The boundary operator of the cell complex K can be split into three summands $\partial = d(\mathcal{M}) + d^a(\mathcal{M}) + d^c(\mathcal{M})$, in the following form:

- the map $d^a(\mathcal{M}) : G(K) \to G(K)$ is defined for a given node σ of a connected $(i, i+1)$-MSF subgraph $\mathcal{M}_k^{(i,i+1)}$ as the sum of summands of $\partial + d(\mathcal{M})$ belonging to $\mathcal{M}_k^{(i,i+1)}$.
- the map $d^c(\mathcal{M}) : G(K) \to G(K)$ is defined as $d^c(\mathcal{M}) = \partial + d(\mathcal{M}) + d^a(\mathcal{M})$. For a node σ of an $(i, i+1)$-MSF subgraph $\mathcal{M}_{i,i+1}$, $d^c(\mathcal{M})(\sigma)$ is the linear combination of the summands of $\partial + d(\mathcal{M})$ which does not belong to $\mathcal{M}_k^{(i,i+1)}$.

The following equalities hold:

(a) $(1 - d(\mathcal{M})\phi(\mathcal{M}))d^a(\mathcal{M})(1 - \phi(\mathcal{M})d(\mathcal{M}))(\tau) = 0$, for every $(i+1)$-cell τ of a $(i, i+1)$ c-tree of $\mathcal{M}$.
(b) $(1 - d(\mathcal{M})\phi(\mathcal{M}))d^c(\mathcal{M})(1 - \phi(\mathcal{M})d(\mathcal{M}))(\tau) = 0$, for every $(i+1)$-cell τ of a $(i, i+1)$ c-tree of $\mathcal{M}$.
(c) Given an i-cell σ of a $(i, i+1)$-MSF subgraph of $\mathcal{M}$, then $(1-d(\mathcal{M})\phi(\mathcal{M}))d^c(\mathcal{M})$ $(1 - \phi(\mathcal{M})d(\mathcal{M}))(\sigma)$ is a linear combination of $(i-1)$-cells belonging to a $(i-1, i)$ c-tree of $\mathcal{M}$.

The map $1 : C_*(K) \to C_*(K)$ is the identity map. The first two equalities are due to the fact that $(1 - \phi(\mathcal{M})d(\mathcal{M}))(\tau) = 0$. The last one is due to the fact that $\mathcal{M}$ is a h-MSF.

An acyclic HSF $\mathcal{H}$ f is gradually constructed from $i = 1$ to k, as follows. Let us suppose that we have constructed $\mathcal{H}$ up to dimension $i-1$. Let σ be an i-cell of an $(i, i+1)$-MSF subgraph of $\mathcal{H}$. Since $d(\mathcal{M})(\sigma) = 0$ and $d^a(\mathcal{M})(\sigma) = 0$, then $\partial(\sigma) = d^c(\mathcal{M})(\sigma)$. If $d^c(\mathcal{M})(\sigma) \neq 0$, then the expression $(1-d(\mathcal{H})\phi(\mathcal{H}))d^c(\mathcal{M})(\sigma)$ is a linear combination $\Sigma_j^r \mu_j$ of critical $(i-1)$-cells μ_j belonging to a $(i-1, i)$-MSF tree.

Two cases can arise:

- The previous linear combination is not zero. Let us choose one of the $(i-1)$-cells μ_t and reverse one directed path $\sigma \to \mu_t = \{\sigma = \gamma_0, \theta_0, \ldots, \gamma_p, \theta_p = \mu_t\}$ created with this algebraic map. The new acyclic MSF $\mathcal{H}$ has the same directed edges than $\mathcal{M}$ excepting those involved in the path $\sigma \to \mu_t$, which are now the directed edges of the path $\mu_t \to \sigma$. Now, the differential arrows are $\{(\sigma_1, \theta_0), \ldots, (\gamma_p, \theta_{p-1})\}$ and the integral arrows are $\{(\theta_0, \sigma), (\theta_1, \gamma_1), \ldots, (\theta_p, \gamma_p)\}$. We also eliminate from the old HSF structure the differential and integral arrows connecting σ to the neighbour cells belonging to its $(i, i+1)$-MSF tree.
- If $(1 - d(\mathcal{M})\phi(\mathcal{M}))d^c(\mathcal{M})(\sigma) = 0$, then σ becomes a critical cell for the new HSF and, in case there is one, the integral arrow (σ, τ) of the old HSF structure connecting σ with the neighbour $(i+1)$-cell τ within its connected $(i, i+1)$-MSF subgraph is replaced by the differential arrow (τ, σ) in the new HSF.

Finally, for a $(i+1)$-cell within a $(i, i+1)$-MSF tree, there are no changes in $\mathcal{H}$. It is straightforward to prove that the MSF structure $\mathcal{H}$ generated in this way satisfies the three conditions for being an HSF for K.

The process for computing a $\mathbb{Z}/2\mathbb{Z}$-HSFs of a k-dimensional cell complex K starting from a h-MSF structure is presented in Algorithm 1. The output of this algorithm is a HSF structure $\mathcal{H} = (V(\mathcal{H}), E(\mathcal{H}))$. We gradually (increasingly in dimension) go over all the i-cells of the complex $\forall i \geq 2$, changing in each step the MSF structure.

Algorithm 1 $\mathcal{HSF}$ computation from h-MSF

Require: A h-MSF $\mathcal{M} = (V(\mathcal{M}), E(\mathcal{M}))$, the boundary operators ∂, $d(\mathcal{M})$, $d^a(\mathcal{M})$ and $d^c(\mathcal{M})$ and the integral operator $\phi(\mathcal{M})$, previously defined.
$\mathcal{H} := \mathcal{M}$
for $i = 2$ to k **do**
  for all i-cell σ **do**
    if σ belongs to a connected $(i, i+1)$-MSF subgraph of $\mathcal{M}$ **then**
      if $(1 - d(\mathcal{H})\phi(\mathcal{H}))d^c(\mathcal{H})(1 - \phi(\mathcal{H})d(\mathcal{H}))(\sigma) == 0$ **then**
        $E(\mathcal{H}) = (E(\mathcal{H}) + \{(\tau, \sigma\}) \setminus \{(\sigma, \tau)\}$, if there is an $(i+1)$-cell τ such that (σ, τ) is an integral arrow of $\mathcal{H}$
      else
        $\Sigma_j^r \mu_j := (1 - d(\mathcal{H})\phi(\mathcal{H}))d^c(\mathcal{H})(1 - \phi(\mathcal{H})d(\mathcal{H})(\sigma)$ where μ_j are critical cells of $\mathcal{H}$
        Let us choose a non-null μ_t $(1 \geq t \geq r)$ and reverse the directed path over $\mathcal{H}$ $\sigma \rightarrow \mu_t = \{\sigma_j = \gamma_0, \theta_0, \ldots, \gamma_p, \theta_p = \mu_t\}$.
        $E(\mathcal{H}) = (E(\mathcal{H}) + \{(\theta_0, \sigma), (\sigma_1, \theta_0), \ldots, (\gamma_p, \theta_{p-1}), (\mu_t, \gamma_p)\} \setminus \{(\sigma, \theta_0), (\theta_0, \gamma_1), \ldots, (\theta_{p-1}, \gamma_p), (\gamma_p, \mu_t\}$.
        $E(\mathcal{H}) = (E(\mathcal{H}) \setminus \{(\sigma, \tau)$ or (τ, σ) / being τ a neighbor cell of σ in a MSF-subgraph
      end if
    end if
  end for
end for

5.1 Sketch of the proof for the algorithm correctness

At zero and one dimension, the spanning forest of the MSF at that level automatically specifies in a natural way that the integral operator $\phi(\mathcal{H}) : C_*(K) \rightarrow C_*(K)$ is a homology integral operator with regards d. In higher dimension, the three conditions needed for being $\mathcal{H}$ an HSF-structure can be easily verified, due to the fact that the integral operator $\phi(\mathcal{H})$ associated to $\mathcal{H}$ is equal to $\phi(\mathcal{M}) + (1 + \phi(\mathcal{M})d(\mathcal{M}))\tilde{\phi}(1 + d(\mathcal{M})\phi(\mathcal{M}))$, where $\tilde{\phi}(\mu_t) = \sigma$, being σ an i-cell within a $(i, i+1)$-MSF subgraph of $\mathcal{M}$ $(i = 1, \ldots, k-1)$ and μ_t defined as previously in this section, and $\tilde{\phi}(\gamma) = 0$ for the rest of cases.

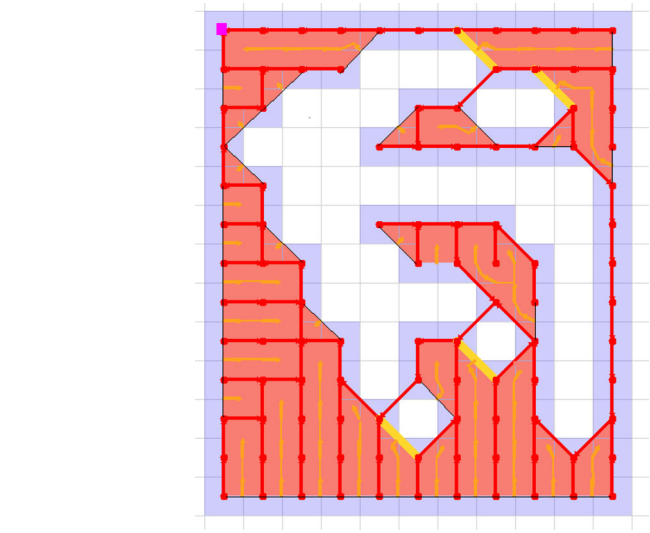
The construction of such a special HSF for a 2D digital object O (and its corresponding cellularization) is shown in Fig. 9. The object O is formed by the set of black (in fact, grey) square pixels. It is obvious that the cell complex $K(O)$ using 8-adjacency between black pixels has one connected component (one connected (0, 1) HSF tree) and four "independent" 1-dimensional holes (specified in this case by four different (1, 2) HSF trees, each of one with only one essential critical 1-cell (marked with yellow thick stroke).

6 Constructing a near-optimal DGVF from an HSF

We can now state the following result concerning near-optimal dgvfs:

By the process of the HSF construction, there is at most one integral arrow coming out from a cell of the cell complex. Let us also note that combinatorial optimality is not always guaranteed. Examples showing the HSF for the Bing's house and one layer of a trabecular bone micro-CT image are shown in Figs. 10 and 11. Figure 10 shows the initial cell complex of the Bing's house, and the integral operator for dimensions

Fig. 9 A 2D example



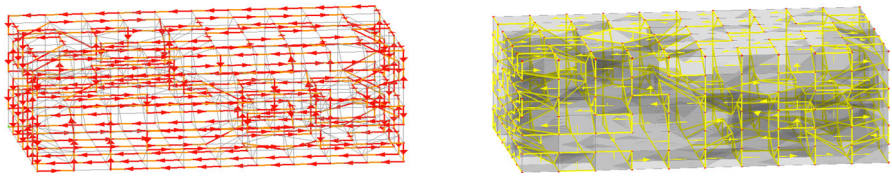


Fig. 10 Bing house

Algorithm 2 dvf computation

Given a h-MSF $\mathcal{M}$ and the canonical boundary operator ∂ on $C_*(K)$, we have constructed an HSF sructure $\mathcal{H}$. In general, a non-necessarily optimal dgvf is determined by the set of integral arrows of the h-MSF. The optimal discrete (non-necessarily gradient) vector field we use in our heuristics is derived from $\mathcal{H}$. Given a connected directed MSF subgraph of $\mathcal{H}$, we simply take as vectors of this dvf all the integral arrows of it.

(0–1) and (1–2). Figure 11 shows the integral operator for dimensions (0–1) and (1–2) and a zoom on a concrete part of the trabecular bone layer.

7 Conclusions

To find a non-redundant global representation of a cell complex by successfully combining combinatorics and algebra is an important step forward that could provide posi-

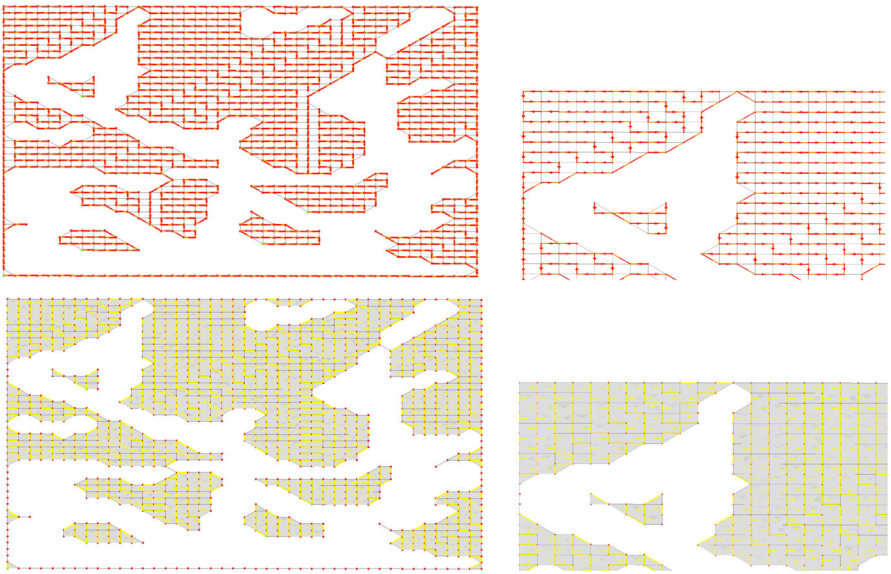


Fig. 11 Trabecular bone

tive answers to several important topological interrogations. The graph-based structure of a Homological Spanning Forest (or HSF, for short) can be thought as an effort to provide a global "skeletal" picture of the cell complex, gluing the connections between cells in a dynamical way.

We propose a refined method based on a direct extraction of the HSF information from the connectivity graph of the cell complex. In order to better understanding, visualizing and checking the accuracy of this heuristic, we place ourselves in the discrete context of digital volumes using coefficients in $\mathbb{Z}/2Z$. Experimental research is done using an HSF software specially designed for this task.

As future work we intend to fathom the analysis capabilities of the HSF notion for the combinatorial optimality problem, working in the applied settings of 3D and 4D digital imagery. It is precisely in fourth dimension, when homology information with integer coefficients explodes in complexity, that is when torsion appears. To adjust the HSF tool to this framework will be a priority in this research.

References

1. Molina-Abril, H., Real, P.: Homological optimality in discrete morse theory through chain homotopies. Pattern Recogn. Lett. **33**(11), 1501–1506 (2012)
2. Whitehead, J.H.C.: Combinatorial homotopy. I. Bull. Am. Math. Soc. **55**, 213–245 (1949)
3. Mari, J., Real, P.: Simplicialization of digital volumes in 26-adjacency: application to topological-analysis. Pattern Recognit. Image Anal **19**(2), 231–238 (2009)
4. Molina-Abril, H., Real, P.: Cell AT-models for digital volumes. GbR 2009, LNCS. **5534**, 314–323 (2009)
5. Munkres, J.: Elements of Algebraic Topology. Addison Wesley, Menlo Park (1984)

6. Delfinado, C., Edelsbrunner, H.: An incremental algorithm for Betti numbers of simplicial complexes on the 3-sphere. Comput. Aided Geom. Des. **12**, 771–784 (1995)
7. Forman, R.: Topology and physics for Raoul Bott. In: Yau, S.T. (ed.) A Discrete Morse Theory for Cell Complexes. International Press, Cambridge (1995)
8. Sergeraert, F.: The computability problem in algebraic topology. Adv. Math. **104**, 1–29 (1994)
9. González-Diaz, R., Medrano, B., Real, P., Sanchez-Pelaez, J.: Algebraic topological analysis of time-sequence of digital images. Lect. Notes Comput. Sci. **3718**, 208–219 (2005)
10. Pilarczyk, P., Real, P.: Computation of cubical homology, (co)homology and (co)homological operations via chain contractions. Adv. Comput. Math. (2014). doi:10.1007/s10444-014-9356-1
11. Joswig, M., Pfetsch, M.E.: Computing optimal morse matchings. SIAM J. Discrete Math. **20**(1), 11–25 (2006)
12. Molina-Abril, H., Real, P.: Homological spanning forest framework for 2d image analysis. Ann. Math. Artif. Intell. **64**(4), 385–409 (2012)
13. Molina-Abril, H., Real, P., Nakamura, A., Klette, R.: Connectivity calculus of fractal polyhedrons. Pattern Recognit. **48**(4), 1146–1156 (2014)
14. Molina-Abril, H., Real, P.: Advanced homological information on 3d digital volumes. In: SSPR 2008, LNCS 5342, pp. 361–371 (2008)
15. Hatcher, A.: Algebraic Topology. Cambridge University Press, Cambridge (2001)
16. Forman, R.: Combinatorial vector fields and dynamical systems. Math. Z. **228**(4), 629–681 (1998)
17. Malgouyres, R., Francés, A.: Determining whether a simplicial 3-complex collapses to a 1-complex is np-complete. In: Discrete Geometry for Computer Imagery. Lecture Notes in Computer Science, vol. 4992. Springer, Berlin, pp. 177–188 (2008)
18. Lewiner, T., Lopes, H., Tavares, G.: Optimal discrete Morse functions for 2-manifolds. Comput. Geom. **26**(3), 221–233 (2003)