Discussion Paper

Helmholtz decomposition for digraphs

Léon Willenborg

7 April 2023

1 Introduction

The present paper[1] is about a concept from analysis (named after the German scientist Helmholtz) that can be translated into a similar concept in graph theory (also named after him). The paper shows in some detail how this translation can be made. The rest of the paper explores this translated concept in the world of graphs. In particular, several concrete examples are considered in the present paper to obtain a better understanding of this concept. It turns out that a method previously proposed by the present author in a totally different context, plays a key role in this translated concept, as is also shown in the present paper.

The classical Helmholtz decomposition applies to vector fields. It is possible to decompose a vector field into two components: one part which is defined by a potential (and hence is rotation free) and another part which is derived from a vector potential (and hence is divergence free). The decomposition is named after Hermann (von) Helmholtz (1821-1894), a German physicist, physician and philosopher, who derived it in [6].[2]

A similar decomposition is possible for digraphs with valuations[3] defined on their arcs and nodes (v-digraphs).[4] It is possible to decompose the valuation in such a network in a similar way: one part that has the property that, for each arc, the sum of the values associated with the arcs on a cycle is zero; the other part has the property that the sum of the values of the ingoing arcs at a node equals the sum of the values of the outgoing arcs, at that node. This means that the notions of cycle and cut are important in the v-digraph setting of the Helmholtz decomposition, as will be shown in the sequel.

We show that this Helmholtz decomposition for v-digraphs can be obtained by the cycle method,[5] which was developed for an entirely different purpose.[6] The cycle method later also proved to be useful in price index theory[7] These can be viewed as examples of applications of the Helmholtz decomposition of which the present author happens to have direct knowledge.

Applications of the Helmholtz decomposition are, however, not considered in the present paper. We only briefly mention some examples here. For a more extensive treatment of some applications see [10]. One application dealt with in [10] is about ranking problems. To determine the ranking of individuals or teams typically tournaments[8] are used, which consists of a cleverly designed system of matches and rules. Each match has a winner and a loser, say, and yields an insight into the relative stength of the contestants involved. As the (imagined) strengths of players or teams are not constants but random variables, one may observe inconsistent tournamant results: if A has won from B and B from C, this does not necessarily imply that A has

1) The author is grateful to Sander Scholtus who reviewed an earlier draft of this paper. His comments resulted in several improvements.
2) Helmholtz, however, was largely anticipated by George Stokes in his paper “On the dynamical theory of diffraction” which was presented and published in 1849 and 1856, respectively. For this and other interesting information on this decomposition see `https://en.wikipedia.org/wiki/Helmholtz_decomposition`.
3) A valuation is a particular kind of function, to be explained later.
4) For earlier work in this area see, for instance, [8], [9] and [10], which also contains applications.
5) The present author later coined this name for the method.
6) The context was land surveying, more in particular leveling, cf. [11].
7) See [12], [13] and [14]
8) In the political or administrative environment one has ballots and polls, which serve a similar purpose of ranking candidates, and in particular, to determine the one who has top rank.

also won from C. If this happens, transitivity is said to be violated, and it is a manifestation of an intransitivity.The Helmholtz decomposition can be used to produce a consistent (transitive) set of results, with as few modifications of the original tournament results as possible.

This problem is comparable to a problem in official statistics, namely price index theory. The problem concerns the modification of a set of intransitive price index numbers[9)] to a set of transitive price index numbers, with minimal modifications. The cycle method can be used to achieve these modifications. About the cycle method and its application to price index theory see [11], [12], [13] and [14].

Another example of an application in statistics of the Helmholtz decomposition in networks is to traffic data, where one may want to gain an insight into the amount of traffic passing through and the amount of local traffic, that is basically only swirling around in the network.

The Helmholtz decomposition for v-digraphs[10)]—the subject of the present paper—is another application of the cycle method. The translation of the Helmholtz decomposition for vector fields to a similar notion on v-digraphs is quite natural, as will be shown, by using certain properties of the original components of a vector field. In [9] another approach is used to obtain this decomposition on v-digraphs, namely one based on the Hodge decomposition, named after W.V.D. Hodge.[11)] In the present paper we use a translation from the original Helmholtz decomposition, which is pretty straightforward and which does not require the more elaborate 'machinery' of the Hodge decomposition.

The paper is organized as follows. In Section 2 we consider the classical Helmholtz decomposition for vector fields. It roughly means that a vector field can be decomposed into two components at every point in its domain: one component can be viewed as a vector field derived from a potential function, whereas the other component can be viewed as the rotation of another vector field. The first field is rotation free, the second field has zero divergence. What this means is explained below. In Section 3 we introduce some key concepts to prepare for the derivations to be made later on in the paper. Digraphs and their underlying graphs are introduced, and also graphs and their orientations. Also spanning trees and spanning ditrees are employed, which are used to compute bases in certain vector spaces, namely cycle spaces and cut spaces. Also the equivalent of the Laplace operator in analysis will be introduced, called the graph Laplacian. It is shown how it is related to the incidence matrix of a digraph or graph, which in turn is closely related to the cut matrix. In Section 4 we review the concepts of cycles and cycle spaces in digraphs on the one hand and cuts and cut spaces in digraphs on the other hand. They are known to be related, as we will elucidate. These concepts are fundamental concepts for the present paper. They are well-known objects in algebraic graph theory. In Section 5 we discuss two methods that can be used to actually compute the Helmholtz decomposition in a v-digraph, namely the cycle method and the cut method. The cycle method was developed by the present author many years ago in a statistical context (land surveying, to be more precise, levelling). Section 6, finally, presents the Helmholtz decomposition for v-digraphs. It is shown how the cycle method or the cut method can be used to compute it for a v-digraph. So, as in the case of the Helmholtz decomposition for vector fields, we not only have proved that the decomposition exists, but we also have presented the methods to compute it. This section is the culmination of

9) In this case transitivity has a somewhat different, but similar, meaning from that in the context of tournaments.
10) The terminology is explained later in the present paper.
11) Hodge inaugurated a new and powerful homology method for analytic and algebraic manifolds, decribed in [7].

the present paper. Section 7 closes the paper with a discussion of the main results and with some ideas for future work, mainly directed at computational issues. The paper is concluded with a list of references.

As to the notation used. This is fairly straightforward. We only mention the use of □ to mark the end of examples or remarks.

2 Helmholtz decomposition for vector fields

The original Helmholtz decomposition applies to vector fields in the area of vector calculus. Our aim here is to give a general description of the decomposition leaving out all technical details needed for an exact treatment of the subject. This is not needed for the motivational purposes we have. However, the interested reader is referred to the literature on the subject for more technical details. We assume here that the reader is acquainted with concepts like 'potential', 'vector potential', 'rotation' and 'divergence'.

Let **F** be a vector field on a bounded domain of space $\mathbb{R}^3$. Under certain fairly general conditions it is possible to express **F** as follows

$$\mathbf{F} = -\nabla\Phi + \nabla \times \mathbf{A}, \tag{1}$$

where Φ is a potential and $\mathbf{A} = (A_x, A_y, A_z)$ is a vector potential in Cartesian coordinates. Using such a coordinate system for Φ as well, we have

$$\nabla\Phi = (\partial_x\Phi, \partial_y\Phi, \partial_z\Phi), \tag{2}$$

which is called the gradient of Φ, where $\partial_x\Phi = \frac{\partial\Phi}{\partial x}$, etc.

Furthermore

$$\nabla \times \mathbf{A} = \begin{vmatrix} \hat{x} & \hat{y} & \hat{z} \\ \partial_x & \partial_y & \partial_z \\ A_x & A_y & A_z \end{vmatrix} = (\partial_y A_z - \partial_z A_y, \partial_z A_x - \partial_x A_z, \partial_x A_y - \partial_y A_x), \tag{3}$$

where $|\cdot|$ denotes a determinant, and $\hat{x}$, $\hat{y}$ and $\hat{z}$ the unit vectors in the x, y and z direction, respectively. $\nabla \times \mathbf{A}$ is called the rotation of **A**. The implicit understanding is that the differentiation in (2) and (3) is allowed. This, in fact, means that Φ and **A** are supposed to be sufficiently smooth functions. Under fairly general conditions we have

$$\nabla \times (\nabla\Phi) = \mathbf{0}, \tag{4}$$

$$\nabla \cdot (\nabla \times \mathbf{A}) = 0. \tag{5}$$

In words, the first identity (4) states that the rotation of the gradient of Φ (see (2)) vanishes. The second identity (5) states that the divergence of the rotation of **A** is zero. The divergence of a vector field **B**, in Cartesian coordinates represented as (B_x, B_y, B_z), is defined as

$$\nabla \cdot \mathbf{B} = \partial_x B_x + \partial_y B_y + \partial_z B_z. \tag{6}$$

Equation (4) holds, provided that

$$\nabla \times (\nabla\Phi) = (\partial_y \partial_z \Phi - \partial_z \partial_y \Phi, \partial_z \partial_x \Phi - \partial_x \partial_z \Phi, \partial_x \partial_y \Phi - \partial_y \partial_x \Phi) = \mathbf{0} \tag{7}$$

holds. This is the case if the order of the partial differentiations can be interchanged, as we shall assume to hold.[12)] If we assume a similar property to hold for the components of **A**, we have

$$\nabla \cdot (\nabla \times \mathbf{A}) = \partial_x(\partial_y A_z - \partial_z A_y) + \partial_y(\partial_z A_x - \partial_x A_z) + \partial_z(\partial_x A_y - \partial_y A_x) = 0. \tag{8}$$

We want to stress that two observations are important, as they can be carried over naturally to the ‘graph theory world’:

1. the component $-\nabla\Phi$ in (1) is derived from a potential function.
2. the component $\nabla \times \mathbf{A}$ in (1) is divergence free.

In Section 6 we discuss how these properties can be carried over to graph theory, and thus how to obtain the equivalent of a Helmholtz decomposition in this area. But in order to do that we need some preparations in Section 3, where we introduce the necessary concepts to be able to state the result.

It is possible to give explicit expressions for the potential Φ and the vector potential **A**. We state them here, for completeness. They are not of any use in the graph theory context, where the two components of the decomposition can also be stated explicitly, as will be shown. We present the results for a special case, namely where the domain of an auxiliary vector field **F**[13)] is $\mathbb{R}^3$ (and hence also that of **A**). Then we have the following representations:

12) In practice this is often the case, so it is hardly a restriction.
13) With the additional technical requirement that $r\mathbf{F} \to \mathbf{0}$ as $r \to \infty$

$$\Phi(\mathbf{r}) = \frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{\nabla' \cdot \mathbf{F}(r')}{|\mathbf{r} - \mathbf{r}'|} dV' \tag{9}$$

$$\mathbf{A}(\mathbf{r}) = \frac{1}{4\pi} \int_{\mathbb{R}^3} \frac{\nabla' \times \mathbf{F}(r')}{|\mathbf{r} - \mathbf{r}'|} dV', \tag{10}$$

where differentiation (via a gradient) with respect to $\mathbf{r}'$ is denoted by ∇' and integration by dV'. In case $\mathbf{F}$ is defined on a bounded domain $\mathbf{V} \subseteq \mathbb{R}^3$ the representations of $\Phi(\mathbf{r})$ and $\mathbf{A}(\mathbf{r})$ are more complex, consisting of two terms each.

3 Preparations

The goal of this section is to introduce the concepts and constructions needed voor the Helmholtz decomposition defined for so-called v-digraphs, which are the graph theoretical counterparts of vector fields. In particular we need the concepts of cycles and cuts in digraphs.

3.1 Underlying graph and orienting graphs

We sometimes need to link digraphs to related graphs, and vice versa. If we have a digraph $\bar{G}$ we sometimes need the underlying graph G. This is the graph with the same node set as G but with arcs (a, b) replaced by edges $\{a, b\}$. Obviously, several digraphs have the same underlying graph. Sometimes we need the reverse operation: we start with a graph and we need to produce a digraph from this structure. This can be done by replacing some edges $\{a, b\}$ by arcs: (a, b) or (b, a). Edges that are not replaced are interpreted as a pair of arcs, as usual.

This operation can be called: edge orientation, resulting in arcs replacing (some) edges. In another interpretation one can view this process as one of selecting arcs: an edge $\{a, b\}$ in fact represents two arcs, (a, b) and (b, a), so edge orientation in fact amounts to selecting one of the arcs (a, b) or (b, a) as a replacement for the edge $\{a, b\} = \{(a, b), (b, a)\}$.

In the present paper we assume the underlying graphs to be connected, meaning that there is a path connecting any two nodes in the graph. A path can be viewed as a sequence of nodes, where each neighboring pair a, b has $\{a, b\}$ is an edge in the underlying graph.

3.2 Spanning trees and ditrees

Let $G = (V, E)$ be a graph, which we assume to be connected. A spanning tree for G is a subgraph $T = (W, F)$ which is a tree, $W = V$ and $F \subseteq E$. So T has the same node set as G, and its edges form a subset of those of G. A cycle is a closed path in which only the first and last nodes are equal.[14] T has no cycles and is connected.

[14] A path in a graph G is a finite sequence of nodes in which successive nodes are adjacent, that is, are joined by an edge in G. In a closed path the initial and final nodes are the same. It is also possible that other nodes on the path are the same. So a cycle is a closed path, but a closed path need not be a cycle.

In case $\bar{G}$ is a digraph, we can also define a concept similar to that of a tree for graphs: a directed tree, for short, a ditree. It is a digraph for which the underlying graph is a tree. So if $\bar{T}$ is a spanning ditree for digraph $\bar{G} = (V, \bar{E})$, it implies that the underlying tree T is a spaning tree for the graph $G = (V, E)$, where the edge set E is derived from the arc set $\bar{E}$ as described above.

3.3 Graph Laplacian

The graph Laplacian[15)] is a graph theoretical translation of the Laplace operator Δ in mathematical analysis, which in cartesian coordinates can be expressed as the following second order partial differential operator (in Cartesian coordinates) in n dimensions:

$$\Delta = \nabla \cdot \nabla = \text{div} \cdot \text{grad} = \frac{\partial^2}{\partial x_1^2} + \cdots + \frac{\partial^2}{\partial x_n^2} \tag{11}$$

where 'div' is the divergence and 'grad' the gradient. For a function $f : \mathbb{R} \to \mathbb{R}^n$ and a function $\Phi : \mathbb{R}^n \to \mathbb{R}$, which satisfy the necessary differentiability criteria, we have

$$\text{div}(f) = \nabla \cdot f = \frac{\partial f}{\partial x_1} + \cdots + \frac{\partial f}{\partial x_n}, \tag{12}$$

$$\text{grad}(\Phi) = \nabla(\Phi) = (\frac{\partial \Phi}{\partial x_1}, \cdots, \frac{\partial \Phi}{\partial x_n}) \tag{13}$$

Intuitively, the divergence measures the flow of f through an infinitesimal cube, whereas the gradient is a vector indicating the local change of Φ.

The graph Laplacian of a graph $G = (V, E)$ is the following matrix

$$\mathcal{L} = \mathfrak{D} - \mathcal{A}, \tag{14}$$

where $\mathfrak{D}$ is the degree matrix and $\mathcal{A}$ is the adjacency matrix of G. The degree matrix is a diagonal matrix where the entry $\mathfrak{D}_{ii}$ is the degree of node i, which is the number of edges incident with i. The adjacency matrix of G, $\mathcal{A}$, is the $n \times n$ (0,1)-matrix such that $\mathcal{A}_{ij} = 1$ if there is an arc (i, j) in G and $\mathcal{A}_{ij} = 0$ if there is no such arc. Here n is the number of nodes in G, i.e. $|V| = n$, where $|\cdot|$ is the function that counts the number of elements in a finite set.

Now we introduce another important matrix for graphs, namely the incidence matrix. Let $G = (V, E)$ be a graph, with $|V| = n$ and $|E| = m$. The incidence matrix $\mathcal{I}$ for G is defined as follows: It is a (0,1)-matrix with $\mathcal{I}_{ij} = 1$ if node i is incident with edge j, and $\mathcal{I}_{ij} = 0$ if this is not the case.

Apart from the adjacency matrix $\mathcal{A}$ for G, the incidence matrix $\mathcal{I}$ is a fundamental matrix that defines G.

15) For more information on graph Laplacians and some of their properties see e.g. [2].

The incidence matrix for the graph G is related to the degree matrix and the adjacency matrix of G, but not quite the graph Laplacian of G itself:

$$\mathcal{I}\mathcal{I}' = \mathfrak{D} + \mathcal{A}, \tag{15}$$

with $\mathfrak{D}$ and $\mathcal{A}$ as defined in (14).

Before we proceed, we to introduce the incidence matrix of a digraph, more specifically of an oriented graph. This is a digraph in which no arc has a counter-arc. Likewise we can say, starting with a graph G that if we replace each edge $\{a, b\}$ by a single arc, either (a, b) or (b, a), we obtain such a digraph $\bar{G}$. We view $\bar{G}$ is an orientation of $G = (V, E)$. Likewise, G is the underlying graph of $\bar{G}$. If G has m edges there are 2^m orientations of G. If (a, b) is an arc in $\bar{G}$, we call b the head and a the tail of this arc. We now define the incidence matrix $\bar{\mathcal{I}}$ of $\bar{G}$. It is a $\{-1, 0, 1\}$-matrix with rows and columns indexed by nodes and arcs of $\bar{G}$. Let (v, e) with $v \in V$ and $e \in \bar{E}$, the arc set of $\bar{G}$. Then

$$\bar{\mathcal{I}}_{ve} = \begin{cases} -1 & \text{if } v = a \\ 0 & \text{if } v \neq a, v \neq b \\ 1 & \text{if } v = b \end{cases} \tag{16}$$

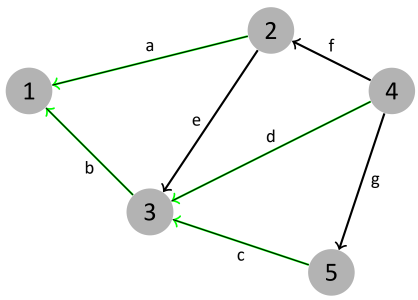

Figure 3.1 A digraph with spanning ditree (arcs coloured green).

Example To illustrate this notion of an incidence matrix for digraphs consider the digraph in Figure 3.1. The corresponding incidence matrix is presented in Table 3.1. □

node	a	b	c	d	e	f	g
1	1	1	0	0	0	0	0
2	-1	0	0	0	-1	1	0
3	0	-1	1	1	1	0	0
4	0	0	0	-1	0	-1	-1
5	0	0	-1	0	0	0	1

Table 3.1 Incidence matrix of the digraph in Figure 3.1.

We can look at the incidence matrix in two ways: column-wise and row-wise. If we look column-wise we basically see how the arcs are defined in terms of the nodes incident to them. Also is indicated, via the sign, if an arc starts at a node (- sign) or arrives at a node (no sign, but actually signifying a + sign). If we look row-wise, we see for each node which arcs depart from it or arrive at it. This is exactly the sort of information we need for our application to the Helmholtz decomposition of v-digraphs. See Section 4.2.

For the incidence matrix $\bar{\mathcal{I}}$ of an oriented graph $\bar{G}$ with underlying graph G the following important identity holds:

$$\bar{\mathcal{I}}\bar{\mathcal{I}}' = \mathfrak{D} - \mathcal{A} = \mathcal{L}, \tag{17}$$

with $\mathcal{L}$ the graph Laplacian of G, as defined in (14).[16] The first equality in (17) holds irrespective of the chosen orientation of G.

We illustrate (17) with an example.

Example In Figure 3.2 a graph G is shown with 4 nodes and 5 edges. In Figure 3.3 a digraph $\bar{G}$ is shown which is an oriented version of the graph G in Figure 3.2.[17]

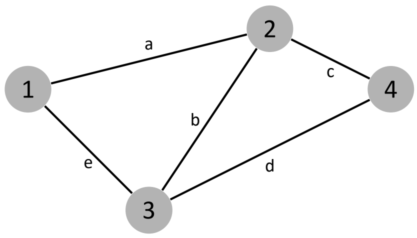

Figure 3.2 Graph G with 4 nodes and 5 edges .

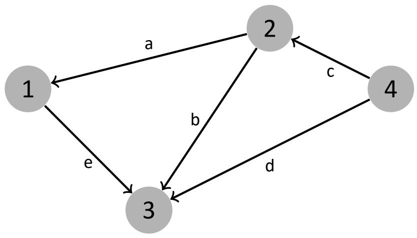

Figure 3.3 Digraph $\bar{G}$ which is an oriented version of the graph G in Figure 3.2.

We now specify the adjacency matrix $\mathcal{A}$, the degree matrix $\mathfrak{D}$ and the incidence matrix $\mathcal{I}$ of G and the incidence matrix $\bar{\mathcal{I}}$ of $\bar{G}$.

$$\mathcal{A} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix}, \tag{18}$$

$$\mathfrak{D} = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}, \tag{19}$$

16) See [4], Lemma 8.3.2 on p. 168.
17) Alternatively expressed: the graph in Figure 3.2 is the underlying graph of the digraph in Figure 3.3.

$$\mathcal{I} = \begin{pmatrix} 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{pmatrix}, \tag{20}$$

$$\bar{\mathcal{I}} = \begin{pmatrix} 1 & 0 & 0 & 0 & -1 \\ -1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & -1 & -1 & 0 \end{pmatrix}. \tag{21}$$

Now we can compute $\mathcal{I}\mathcal{I}'$ and $\bar{\mathcal{I}}\bar{\mathcal{I}}'$ and express the results in terms of $\mathcal{A}$ and $\mathfrak{D}$, as given in (18) and (19), respectively:

$$\mathcal{I}\mathcal{I}' = \begin{pmatrix} 2 & 1 & 1 & 0 \\ 1 & 3 & 1 & 1 \\ 1 & 1 & 3 & 1 \\ 0 & 1 & 1 & 2 \end{pmatrix} = \mathfrak{D} + \mathcal{A}, \tag{22}$$

$$\bar{\mathcal{I}}\bar{\mathcal{I}}' = \begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ 0 & -1 & -1 & 2 \end{pmatrix} = \mathfrak{D} - \mathcal{A} = \mathcal{L}, \tag{23}$$

which is the graph Laplacian of $\bar{G}$.

Suppose that another orientation of G would have been chosen, for instance the one in Figure 3.4.

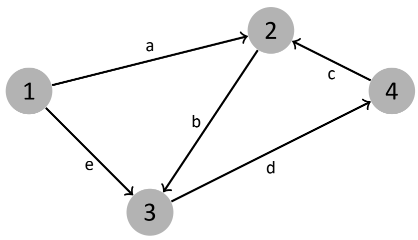

Figure 3.4 Digraph $\bar{\bar{G}}$ which is another oriented version of the graph G in Figure 3.2.

The incidence matrix $\bar{\bar{\mathcal{I}}}$ for the digraph in Figure 3.4 is

$$\bar{\bar{\mathcal{I}}} = \begin{pmatrix} -1 & 0 & 0 & 0 & -1 \\ 1 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 \\ 0 & 0 & -1 & 1 & 0 \end{pmatrix}. \tag{24}$$

We have

$$\bar{\bar{\mathcal{I}}}\bar{\bar{\mathcal{I}}}' = \begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ 0 & -1 & -1 & 2 \end{pmatrix} = \mathfrak{D} - \mathcal{A} = \mathcal{L}, \tag{25}$$

which is the same result as (23), showing that the difference in orientation of G does not affect the result.

Suppose that the values x_1, x_2, x_3, x_4 are associated with the nodes 1, 2, 3 and 4, respectively.[18] Write $x = (x_1, x_2, x_3, x_4)'$. Then

$$\begin{aligned} x'\mathcal{L}x &= x'\bar{\mathcal{I}}\bar{\mathcal{I}}'x = (\bar{\mathcal{I}}'x)'(\bar{\mathcal{I}}'x) \\ &= (x_1 - x_2)^2 + (x_1 - x_3)^2 + (x_2 - x_3)^2 + (x_2 - x_4)^2 + (x_3 - x_4)^2, \end{aligned} \tag{26}$$

where $\mathcal{L}$ is as in (23) and $\bar{\mathcal{I}}$ as in (21).[19] (26) can be interpreted as an elastic energy of a network, as defined in [15]. □

Remark The graph Laplacian is related to the cut method and (hence) also to the Helmholtz decomposition in v-digraphs, as is explained in Section 5.3. This is the reason why the graph Laplacian is discussed in the present section. □

4 Cycles and cuts

4.1 Cycles

In Section 3.2 we already encountered to notion of cycle in graphs, when we introduced trees and ditrees. However, there is more to say about this notion. It is in fact a key notion for the present paper. A large portion of the section can be explained for graphs, but is applicable to digraphs as well. This is what we need for the Helmholtz decomposition.

4.1.1 Cycle space

Let G be a graph, which is not a tree, so it has cycles. A cycle in G is a connected subgraph of G such that each of its nodes has degree 2.

First we remark dat it is possible to define an algebraic structure on the set of cycles in G . This allows us to combine ('add' or 'sum') two cycles, and the result is another cycle. The sum $\oplus$ is defined in such a way that adding two cycles with a common edge, the result is a combined cycle with the common edge removed. Symbolically, $e \oplus e = 0$, for any edge e of G.

18) Put differently, they define a valuation on the node set of G.

19) Instead of $\bar{\mathcal{I}}$ we could have taken $\bar{\bar{\mathcal{I}}}$ just as well. This shows that the orientation of two digraphs with the same underlying graph yields the same result.

Example In Figure 4.1 two cycles in a graph are shown and their sum. The cycles have a common edge, b, which is eliminated as a result of the addition, yielding a new cycle. The remaining edges, which are unique in the two cycles, are maintained and therefore appear in the resulting cycle. ☐

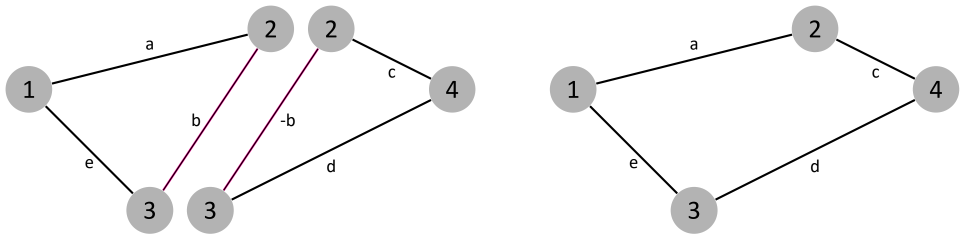

Figure 4.1 Adding two (non-oriented) cycles in a graph: $\{a, b, e\} \oplus \{b, c, d\} = \{a, c, d, e\}$**.**

It can be shown that the cycles in a graph form a vector space: one can add or subtract cycles and get cycles as a result. The scalar field is not $\mathbb{R}$ or $\mathbb{C}$ as usual, but the finite field $\mathbb{F}_2$ consisting only of the elements 0 and 1 and with two operations addition $\oplus$ and multiplication $\cdot$, defined as follows:

$$
\begin{aligned}
&0 \oplus 0 = 1 \oplus 1 = 0, \\
&0 \oplus 1 = 1 \oplus 0 = 1.
\end{aligned}
\tag{27}
$$

There is no need to define multiplication, as we do not need it in our applications.

So the cycle space of a graph G is a vector space over $\mathbb{F}_2$. As more well-known vector spaces it has a basis consisting of base cycles, or elementary cycles as we prefer to call them. These are a subset of all the cycles of G that can be used to generate any of the cycles in G. In fact, there are typically more bases to the cycle space of a graph, as is the case in 'ordinary' vector spaces over $\mathbb{R}$ or $\mathbb{C}$. Computing a cycle basis, or equivalently, a set of elementary cycles, of a graph G can be done with the help of a spanning tree. This is further explained in Section 4.1.2.

Now we turn to cycle addition in digraphs which is a bit different from that in graphs. The following example illustrates it.

Example Consider the digraph in Figure 3.4. Now we are dealing with arcs (which are directed) rather than with edges (which are not). Now it may be necessary to use a reversed arc. If an arc has a label, say b, then we use the label $-b$ to indicate the counter arc; the minus sign in the label indicates opposite orientation. So in Figure 3.4 b represents the arc $(2, 3)$ and $-b$ the counter-arc $(3, 2)$, with reversed orientation. Now the idea is that if we have a cycle we orient it, so that we can traverse it in the direction of its orientation, starting at some node and ending there as well, after having traversed the cycle. So if we consider the cycle containing the nodes 1, 2 and 3 we may represent the cycle as either $\{a, b, -e\}$ or as $\{e, -b, -a\}$. The order of the arcs in each set is irrelevant (as they are sets). If we consider a second cycle, for instance the one containing the nodes 2, 3 and 4, we have either the oriented cycle $\{b, d, c\}$ or its reversely oriented counterpart $\{-b, -c, -d\}$. We assume that we can combine two (oriented) cycles if the underlying (non-oriented) cycles have edges in common, and, viewed as a pair of arcs, one arc belongs to one of the (oriented) cycle and its counter-arc to the other (oriented) cycle. [20] For

[20] If an arc is indicated by a label a its counter-arc has label $-a$.

instance, for the two cycles just indicated, if the first cycle is represented as $\{a, b, -e\}$ and the other as $\{-b, -c, -d\}$, this is the case: arcs b and $-b$ cancel each other, and what is left is the set $\{a, -e, -c, -d\}$, which is indeed also a cycle traversed in a particular way. Symbolically we write the combination of the two (oriented) cycles as a sum

$$\{a, -e, -c, -d\} = \{a, b, -e\} + \{-b, -c, -d\} \tag{28}$$

If we take the opposite orientation of the first cycle $\{-a, -b, e\}$ and the opposite orientation of the second one $\{b, c, d\}$, we can again combine them to a cycle $\{-a, e, d, c\}$, which, of course, is the resulting cycle we found previously, but oriented in the opposite direction. Symbolically we can write this combination of two (oriented) cycles as

$$\{-a, e, d, c\} = \{-a, -b, e\} + \{b, c, d\} \tag{29}$$

Obviously, the orientations of both cycles can be incompatible, in the sense that no pair of arc and counter-arc occurs in the two cycles, and no new cycle can be formed. The addition of such incompatible (oriented) cycles is not defined.

The addition ('+') of oriented cycles is an operation that is commutative ($a + b = b + a$) and associative ($(a + b) + c = a + (b + c)$). Adding a cycle d and its counterpart $-d$ (with opposite orientation) yields the 0-cycle which is actually the empty set ($d + (-d) = 0$). □

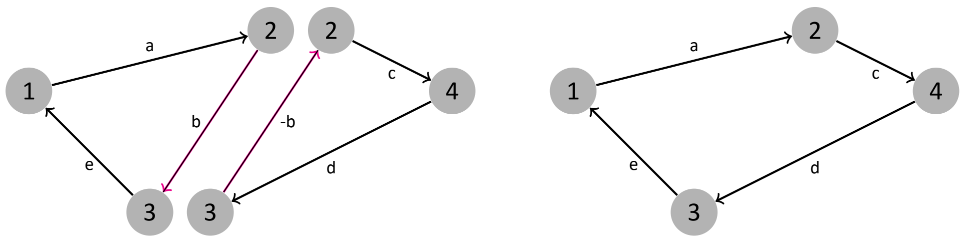

Figure 4.2 Adding two (oriented) cycles in a digraph: $\{a, b, e\} + \{-b, c, d\} = \{a, c, d, e\}$.

The labels associated with the arcs of the digraphs should be thought of as numbers (say elements of $\mathbb{R}$) instead of as symbolic labels as they were in the case of graphs. That is how they will be used later on in this paper, when v-digraphs are introduced and used.

4.1.2 Computing elementary cycles

We consider an example to show how to compute a cycle space. Consider the digraph in Figure 3.1, in which also a spanning tree is indicated (green arcs). We use the spanning tree to compute a basis of the cycle space, which is a set of elementary cycles. We explain below how the spanning tree is used for this purpose.

The arcs not situated on the spanning tree in Figure 3.1, that is, e, f and g, are coloured black. They are important for the generation of a set of elementary cycles: each one of those arcs generates such a cycle. We start with the spanning tree and then consecutively add each of the black arcs, e, f and g. So we start with arc e and add it to the spanning tree. This yields

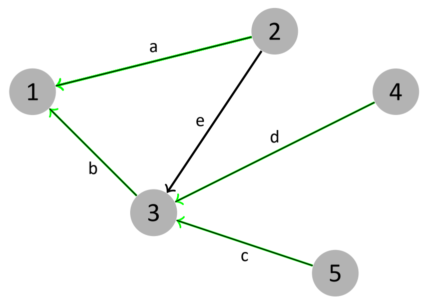

Figure 4.3 Arc e is added, generating elementary cycle $\{a, -b, -e\}$.

elementary cycle $\{a, -b, -e\}$. See Figure 4.3. The arcs c and d as well as the nodes 4 and 5 should be discarded, as they are not part of the elementary cycle.

For the next elementary cycle we add arc f to the spanning tree. This yields elementary cycle $\{a, -b, -d, f\}$. See Figure 4.4. Arc c and node 5 are not part of this elementary cycle and should therefore be discarded.

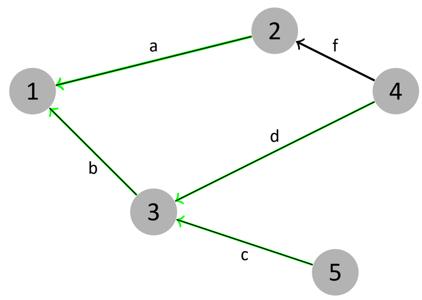

Figure 4.4 Arc f is added, generating elementary cycle $\{a, -b, -d, f\}$.

The last arc to add to the spanning tree is arc g. It yields elementary cycle $\{-c, -g, d\}$. See Figure 4.5. In this case the arcs a and b as well as nodes 1 and 2 should be discarded as not being part of the elementary cycle.

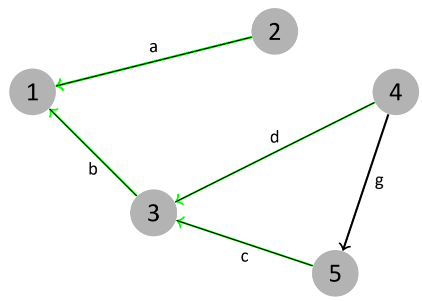

Figure 4.5 Arc g is added, generating elementary cycle $\{-c, -g, d\}$.

In this case we have generated a cycle basis using the spanning tree with arcs a, b, c, d. A different choice of spanning tree would have yielded another basis of elementary cycles. As in more familiar vector spaces, like $\mathbb{R}^2$ or $\mathbb{R}^3$, a basis is not unique.[21)] You can pick any that you want: they all span the same cycle space, which is what we need in the present paper.[22)]

21) Provided there are sufficient cycles in the digraph at hand, of course.
22) So theoretically all these cycle space bases are equivalent. Computationally it is perhaps another matter. A particular basis may be more attractive than another one, especially for big digraphs.

4.1.3 Cycle matrix

We can describe the set of elementary cycles for a digraph $G = (V, E)$ with the help of a cycle matrix. This is a (-1,0,1)-matrix of order $c \times m$, where c is the number of elementary cycles in G and m the number of edges in G. To illustrate this consider the digraph in Figure 3.1. It has three elementary cycles as we have seen in Section 4.1.2, with the following cycle matrix:

$$C = \begin{pmatrix} 1 & -1 & 0 & 0 & -1 & 0 & 0 \\ 1 & -1 & 0 & -1 & 0 & 1 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & -1 \end{pmatrix}. \tag{30}$$

Remark In case G is a graph it should be noted that the cycle matrix is a (0,1)-matrix: we only need two values to indicate whether an edge is present (value= 1) or absent (value = 0). It should be remembered that the field to be used is $\mathbb{F}^2$, not $\mathbb{R}$. But as for the present paper, we need digraphs and not graphs, so we need not to worry about this case. □

4.1.4 Changing cycle basis

The row vectors of a cycle matrix for a digraph G span the cycle space. Typically there are several bases that span this space. If we start with another spanning for G we find another basis. Suppose we have two cycle matrices C_1 and C_2 for G. It is possible to express each elementary cycle for the second basis in terms of the elementary cycles in the first basis using the $\oplus$ operation (and vice versa). This amounts to

$$C_2 = \Omega C_1, \tag{31}$$

where Ω is a $c \times c$ invertible matrix, where c are the number of elementary cycle = dimension of the cycle space of G.

4.2 Cuts

Intuitively a cut in a digraph is about partitioning the nodes into two (disjunctive) sets and consider the arcs incident to nodes from one set and nodes from the other set. These nodes partitioning sets are called shores.

Remark In the area of Operations Research there is a well-known theorem about cuts and flows in a flow network: max flow = min cut. This means that the maximum flow through a flow network cannot exceed the total capacity of the most restrictive bottleneck of the flow network. This is its minimum cut. □

As in the case of cycles we can talk of a vector space called the cut space, with a basis consisting of elementary cuts that span the cut space. In fact, typically, a cut space has several bases. As in the case of cycle space spanning trees can be used to compute a basis for the cut space.[23)]

23) See [4], Section 14.1.

However, there is a simple way to find a basis for the cut space of a graph or digraph, namely via the incidence matrix, which we shall use.

Now consider the incidence matrix in Table 3.1. Note that the sum of its rows is the 0-vector. Hence the row rank is less than 5. In fact it is 4, as some manipulation of the rows shows. So if we leave out one row, say the last one, the remaining rows are independent. We then obtain the cut matrix D in (32). The rows of D form a basis of the cut space of the digraph in Figure 3.1.

$$D = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & -1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & -1 & -1 \end{pmatrix} \tag{32}$$

We want to illustrate the basis of the cut space defined by the rows in matrix D in (32), induced by nodes $1, 2, 3$ and 4 of the digraph in Figure 3.1. We start with node 1, and then consider nodes 2, 3 and 4 consecutively. In each case we consider the cut set generated. The results are shown in Figures 4.6, 4.7, 4.8 and 4.9, respectively.

It should be noted that this is not the only way to find a basis for the cut space. It can also usually done with the help of a spanning tree, as is illustrated in Section 4.2.1 by an example.

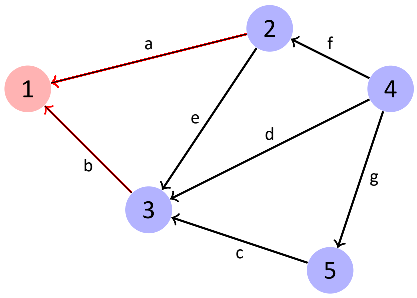

Figure 4.6 Red and blue shores induced by node 1. The cut set is $\{a, b\}$.

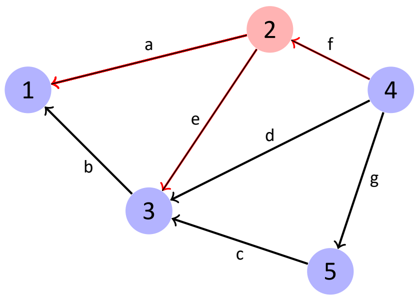

Figure 4.7 Red and blue shores induced by node 2. The cut set is $\{-a, -e, f\}$.

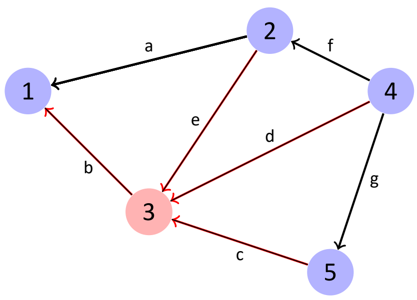

Figure 4.8 Red and blue shores induced by node 3. The cut set is $\{-b, c, d, e\}$.

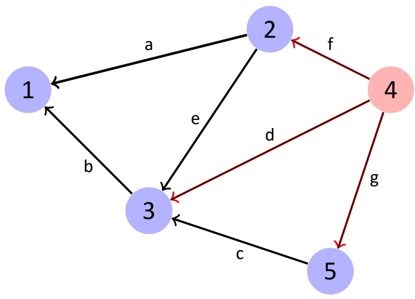

Figure 4.9 Red and blue shores induced by node 4. The cut set is $\{-d, -f, -g\}$.

4.2.1 Using a spanning ditree to find a basis of the cut space

As in the case of finding a cycle basis, we use a spanning ditree to find a basis for the cut space of a digraph. So let $G = (V, E)$ be a digraph, and let T be a spanning ditree of G. Let $V = V_1 \cup V_2$ be a partition of V. This means that $V_1 \neq \emptyset$ and $V_2 \neq \emptyset$ and $V_1 \cap V_2 = \emptyset$. V_1 and V_2 are called shores. The arcs in E can be partitioned into two groups: arcs connecting nodes in the same shore and arcs connecting nodes in different shores. The latter type of arcs have our interest in defining cut sets.

Now we use a spanning ditree to find a basis of the cut space. We illustrate this with the graph in Figure 3.1, where the arcs a, b, c and d define the spanning ditree. We consider each of these arcs separately and in each case we color the head node red and the tail node blue. We also color the nodes at the same side of the spanning ditree as the head node red. Similarly, the nodes at the same side of the spanning ditree as the tail node are colored blue. Then we obtain two shores, the blue shore and the red shore. Then we consider the arcs connecting nodes on different shores.

We start with arc a. In Figure 4.10. Node 1 is the head node of a and hence is colored red and 2 is the tail node of a and is therefore colored blue. Nodes 3, 4 and 5 are on the same side of the spanning ditree as the head node and therefore are all colored red. Node 2 is the only node in the other shore. So the arcs with nodes in different shores are a, e and f. As f is pointing from the red shore to the blue one, we take its counter-arc $-f$ to add to the cut set, which in this case is $\{a, e, -f\}$. The next arc on the spanning tree is b. In this case the red and blue shores as well as the resulting cut set $\{b. - e, f\}$ are shown in Figure 4.11. Then we have arc c. The red and blue shores as well as the cut set in this case is $\{c, -g\}$ are shown in Figure 4.12, where we took the counter-arc $-g$ of g so to have an arc pointing from the blue to the read shore. The final arc on the spanning tree to consider is d. Figure 4.13 shows the two shores as well as the cut set $\{d, f, g\}$ that result in this case.

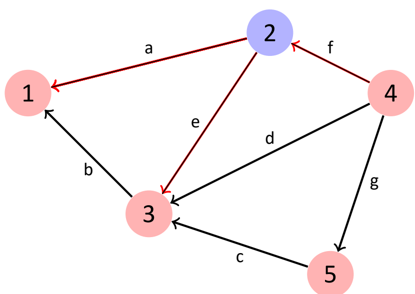

Figure 4.10 Red and blue shores induced by arc a in the spanning tree. The cut set is $\{a, e, -f\}$.

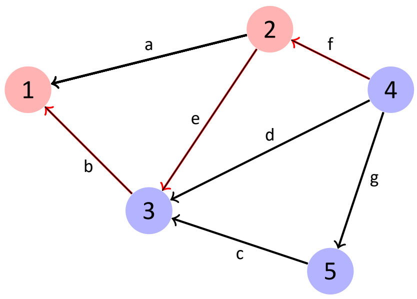

Figure 4.11 Red and blue shores induced by arc b in the spanning tree. The cut set is $\{b, -e, f\}$.

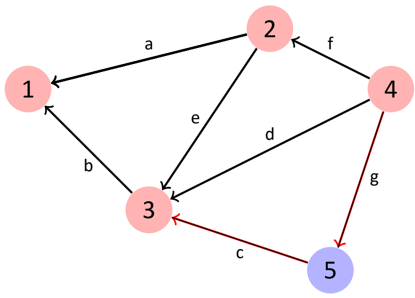

Figure 4.12 Red and blue shores induced by arc c in the spanning tree. The cut set is $\{c, -g\}$.

So the cut sets generated by the arcs situated on the spanning tree in Figure 3.1 lead to a representation of the cut matrix D shown in Table 4.1.

It should be noted that any spanning tree of the digraph in Figure 3.1 has all but three arcs, as it has three elementary cycles: for each elementary cycle 1 arc is missing. This implies that in the cut matrix associated with a particular spanning tree the columns corresponding to these 'left out' arcs have entries equal to 0.

4.3 The relationship between cycles and cuts

Let D be a cut matrix of some digraph G. The cycle space is formally defined as the orthogonal complement of D, that is all $(-1, 0, 1)$-vectors c such that $Dc = 0$. It is easy to see that c is actually a cycle. This is the case because for each arc which enters a node (or exits it) there should be another arc, exiting this node (or entering it). This is precisely what a cycle 'does' for each of the nodes that are part of it. So if C is a cycle matrix for G and D a cut matrix for G, we have

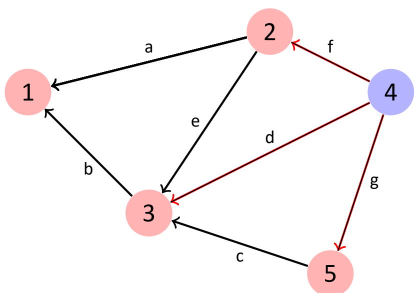

Figure 4.13 Red and blue shores induced by arc d in the spanning tree. The cut set is $\{d, f, g\}$.

no.	a	b	c	d	e	f	g
1	1	0	0	0	1	-1	0
2	0	1	0	0	-1	1	0
3	0	0	1	0	0	0	-1
4	0	0	0	1	0	1	1

Table 4.1 Cut matrix derived from the spanning tree in Figure 3.1.

$$DC' = 0. \tag{33}$$

Actually, (33) is a very important identity, with crucial consequences. For instance, it says that the vectors associated with the elementary cycles in C, spanning the cycle space, are orthogonal to the vectors associated with the elementary cuts, spanning the cut space. So consequently, the cycle space and the cut space of a (di)graph are orthogonal complements, for short orthoplements. Together they span $\mathbb{R}^m$, where m is the number of arcs in the digraph. Summarizing, the following properties hold:

$$\begin{aligned} &\text{cycle space of } G \perp \text{cut space of } G \\ \mathbb{R}^m &= \text{cycle space of } G \oplus \text{cut space of } G, \\ c &= \text{dim cycle space}, \\ m - c &= \text{dim cut space}, \end{aligned}$$

where '$\oplus$' is the direct sum of vector spaces, and 'dim' is an abbreviation of 'dimension' (of a vector space). It is assumed that the underlying graph of $G = (V, E)$ is connected, and $|E| = m$. From the above, we can also conclude that the decomposition of $\mathbb{R}^m$ for a digraph G is unique.[24)]

Example To illustrate (33) we consider the matrices C and D given in (30) and (32) that both belong to the digraph in Figure 3.1.

$$\begin{aligned} DC' &= \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & -1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & -1 & -1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 1 \\ -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \end{aligned} \tag{34}$$

as should be the case, as a specific case of a general property. □

24) Anticipating on the sequel, it also implies that the Helmholtz decomposition for a given v-digraph is unique.

Equality (33) is important for the sequel, i.c. the cycle method and the cut method. It is at the heart of the Helmholtz method for v-digraphs, as will be shown.

5 Cycle method and cut method

In the present section we discuss two methods that can be used to rectify valuations on digraphs: the cycle method and the cut method. The cycle method makes sure that the adjusted valuation has the property that if the adjusted valuations are summed over any cycle, the result is 0. For the adjustment of the valuations by the cut method it holds at each node, the sum of what goes into the node also comes out; there is no accumulation at the node.

The cycle method was derived in a statistical context.[25)] Using similar ideas that led to the cycle method, the cut method was actually derived for the first time in the present paper. Both methods are explained below in separate subsections.

As it turned out both the cycle method and the cut method can be used to achieve the Helmholtz decomposition. But that will be explained in Section 6.

5.1 v-digraphs

In the preparations (Section 3) we only looked at graphs and digraphs. They were all that were needed to discuss, cycles, cuts, etc. But now we want to look at richer structures, namely digraphs with valuations on arcs and nodes, for short v-digraphs.[26)] More formally a v-digraph is a quadruple $G = (V, E, f_V, f_E)$, where

- $G = (V, E)$ a digraph,
- $f_V : V \to \mathbb{R}$ a function, a valuation defined on the set of nodes of G,
- $f_E : E \to \mathbb{R}$ a function, a valuation defined on the set of arcs of G,
- f_V and f_E are linked through a connecting equation, which we explain below.

We are not interested in developing a general theory about v-digraphs and admissable connecting equations, as our goal is much more limited in the present paper. We only need connecting equations of the type

$$f_E(v, w) = f_V(w) - f_V(v), \tag{35}$$

if $(v, w) \in E$. Note that if arc (w, v) is added, we can extend f_E by defining

25) In an area of land surveying, called levelling. The problem is to adjust measured height differences in such a way that the height difference between any two points in the measuring network is independent of the way it is computed, that is by summing the adjusted height differences for the arcs on a path connecting the two points. See [11], [12] for additional background information. Later is was applied in the context of price indices, to derive transitive price index values from nontransitive ones. See [13] or [14].

26) Valuations are real-valued functions on the set of arcs and nodes, which are mutualy related in a specific way.

$$f_E(w, v) = -f_E(v, w), \tag{36}$$

In the present paper we are only interested in connecting equations such as (35), but it should be borne in mind that other choices for connecting equations are of use, such as the following example presents.

Example In the context of price indices (see for example [12], [13] and [14]) the following connecting equation is used:

$$f_E(v, w) = f_V(w)/f_V(v), \tag{37}$$

where $f_V(x) > 0$ for all $x \in V$. (37) can also be expressed as

$$\log f_E(v, w) = \log f_V(w) - \log f_V(v), \tag{38}$$

which is more convenient in computations using the cycle method. The computations aim at adjusting price index values computed with a nontransitive price index formula to obtain values that are transitive. See the references cited if one should be interested in this application. ☐

Note that if f_V has been defined, f_E can be derived using the connecting equation, like (35), (37) or (38). But not the other way round. In fact, it may even be the case the f_E is defined in such a way that no f_V exists. This is the case when it is not true that for every cycle the sum of the values of the arcs in this cycle is 0, which we call the 0-cycle condition. It may be possible to adjust the values of f_E such that the 0-cycle condition holds for the adjusted values. In fact, that was the aim of the cycle method (see Section 5.2). And even in case the f_E form a consistent set (in the sense that they satisfy the 0-cycle condition), there is no unique f_V that is implied. It is easy to see that for any f_V that satisfies (35), (37) or (38), so does $f_E + h$ for any constant $h \in \mathbb{R}$, or hf_E for $h > 0$. This is a gauge property, which one also encounters in physics, for example in electromagnetism and in the classical theory of gravity.

It is only in case f_V and f_E are consistent that it is possible to define a v-digraph based on these valuations. However, it is possible to start with a valuation f_E defined on the arcs of G, which is not consistent, in the sense that the 0-cycle property does not hold. This is equivalent to the existence of a pair $(v, w) \in V \times V$ such that no derived value of f_E exists, which is independent of the path connecting v to w. Given a path γ in G connecting v to w, viewed as a set / sequence of arcs, we can define

$$f_E^{\gamma}(v, w) = \sum_{e \in \gamma} f_E(e). \tag{39}$$

If this value is independent of path γ we can define

$$f_E(v,w) = \sum_{e\in\gamma} f_E(e), \tag{40}$$

for any path γ connecting v to w. In case the 0-cycle property does not hold, there is a pair of nodes (v^*, w^*) and two paths γ_1, γ_2 connecting v^* to w^*, such that $f_E^{\gamma_1}(v^*, w^*) \neq f_E^{\gamma_2}(v^*, w^*)$. In this case one cannot define an $f_E(v^*, w^*)$.

We call a digraph with a triple $G = (V, E, f_E)$, such that

- $G = (V, E)$ a digraph,
- $f_E : E \rightarrow \mathbb{R}$ a function (valuation), not consistent, in the sense that the 0-cycle property does not hold,

a pre-v-digraph. This means that the valuation of two nodes (v, w) is not defined, as the result depends of the path in G connecting v and w. The cycle method (see Secion 5.2) can be defined to produce a v-digraph from a pre v-digraph. In [11] this is explained.

To present a concrete example of a v-digraph and its valuations, consider the following example.

Example We consider an example from land surveying, i.e. levelling. The problem is to measure the height difference between two points at the earth's surface. Let $G = (V, E)$ be a leveling network, which is a digraph, where the set of nodes correspond to points from which height (difference) measurements have been made and E, the set of arcs corresponds to height difference measurements.[27)] Let $a = (a_1, a_2)$ be an arrow. Its tail, a_1, indicates from which point a height (difference) measurement was made. Its head a_2 indicates which point was the target of the measurement. The valuation f_E on E are measured height differences. For arrow a this would be $f_E(a)$. So far we only dealt with height differences. But if the height of a selected node is given, say at sea level, then the heights of the nodes in V could be given, provided the height differences are consistent. This means that the height between two nodes should be independent of the way two nodes in G are connected by a path which is a sequence of arcs, for which height difference are associated. Typically this is not the case with measured height differences. But applying the cycle method to the measured height differences yields a consistent set of adjusted height differences. That was precisely the aim of the cycle method (see Section 5.2). This allows one to associate a height to each of the nodes v in V: for node v this would be $f_V(v)$. The height differences and heights are connected: for an arc $a = (a_1, a_2)$ we have $f_E(a) = f_V(a_2) - f_V(a_1)$. This connection between the valuations f_V and f_E shows that only height differences matter not absolute heights: if the node chosen initially at sea level would in fact be at a height h, we simply need to add this value to the heights of the remaining nodes to get a consistent set of heights, that is in agreement with the (measured and adjusted height differences) and the height of a particular node in the network.[28)] □

It may be tempting to think of a valuation as a flow and the v-digraph as a flow network. But this is very misleading—if not plain wrong—as there are major differences. A flow network has a source (a node with only outflow) and a sink (a node with only inflow), the flows on each arc may be restricted in (absolute) size by a capacity (or upper bound), the inflow and outflow at each

27) Or averages thereof in case multiple measurements were made for an arc.
28) This addresses the issue of gauge invariance in physics, which also applies to potentials.

node are equal (except for the source and the sink). For a v-digraph such restrictions do not necessarily hold, alhough it may possess some of these properties.

5.2 Cycle method

In this section we discuss the cycle method briefly. Only results will be given, no derivations, as these can be found elsewhere.

We start with a pre v-digraph (V, E, f_E), where we shall use x instead of f_E. V is a (finite) set of nodes, E a (finite) set of arcs ('directed edges'), and $x : E \rightarrow \mathbb{R}$ a valuation. A concrete example of such a pre v-digraph is the following one.

Example We consider a land surveying application, or more specifically, a levelling application, which aims at measuring the height difference between two points on the surface of the earth. The distance between these points is not so large that the curvature of the earth has to be taken into account, but also big enough so that the height difference between these points cannnot be measured directly. Instead intermediate points have to be used to measure their height differences. Using these intermediate measurements, the idea is that the height difference between the target points can be determined. But unfortunately, the intermediate height difference measurements are afflicted with measurement errors. The network $G = (V, E)$ consists of intermediate points (elements of V) which are reference points involved in measurements of height differences. An arc $(v, w) \in E$ indicates that the height difference between v and w has been measured, and that the measurement was made from the point v. □

A second example of the use of the cycle method is quite different, in the sense that the method does not correct measurement errors but adjusts particular kinds of imperfections in certain estimates (of price index numbers).

Example In case of price indices a problem may exist which is not due to measurement errors in the data, but stems from the fact that the price index formula used is not transitive. For a transitive price index $P^{i,j}$ comparing the base period i with a reference period j it holds that, if k is another period, it holds that $P^{i,k} = P^{i,j}P^{j,k}$. So for a nontransitive (or intransitive) price index, there are periods i, j, k such that $P^{i,k} \neq P^{i,j}P^{j,k}$. This in fact shows that direct comparison on periods i and j does not yield the same result as comparing i and j indirectly, via an intermediate period k. This in fact means that the comparison of periods i and j depends on the path to link period i to period j. It holds more generally, with paths consisting of several intermediate periods. Price indices that are not transitive are also said to show drift.[29] So transitivity is a highly desirable property of price indices. However, not all price indices are transitive. Even well-known price indices such as those named after Paasche, Laspeyres and Fisher are not transitive. □

The adjustment method employed to correct for measurement errors (as in case of land surveying) or to 'transitivize' intransitive price index numbers is the cycle method.[30] The cycle

29) When comparing the price levels in different countries, one may have the same problem. In fact Hill's method (see [1], p. 256) tackles this problem by specifying for each pair of countries a single path which should be used to compare them. This does indeed avoid inconsistencies, but these is no substantive reason why only these particular paths should be used.

30) It should be noted that the notation used here is slightly different from that used in [12], [13] or [14].

method uses a weight matrix W, defined on the arcs of the digraph G[31] that the user can use to control the degree of adjustment applied to the original values. However, in the applications of the cycle method to the Helmholtz decomposition in v-digraphs, W is chosen to be an identity matrix (of the appropriate order). The matrix C is the cycle matrix, defined in Section 4.1.3. We have

$$\hat{x} = x - WC'(CWC')^{-1}Cx = P_W x, \tag{41}$$

where

$$P_W = I - WC'(CWC')^{-1}C. \tag{42}$$

We put

$$Q_W = WC'(CWC')^{-1}C. \tag{43}$$

Both P_W and Q_W are projections, or equivalently, symmetric and idempotent matrices, that is, $P_W' = P_W$, $Q_W' = Q_W$ and $P_W^2 = P_W$, $Q_W^2 = Q_W$.

Remark In the Helmholtz decomposition in v-digraphs $W = I$, where I is the identity matrix of the appropriate order. Taking $W = I$ is crucial for the Helmholtz decomposition for v-digraphs. In fact, if we would use a $W \neq I$ chances are that such a decomposition does not exist, as the condition $CWD' = 0$ would have to be satisfied. This, however, is not likely to be the case since in the cycle method W is a matrix based on subjective choices of a user of the method. In contrast C and D are 'objective' matrices, derived from the v-digraph. So in the context of the Helmholtz decomposition $W = I$ is the only option for W . □

So with $W = I$, with I the $m \times m$ identity matrix, we have:

$$P = I - C'(CC')^{-1}C, \tag{44}$$
$$Q = C'(CC')^{-1}C. \tag{45}$$

where, we have written P and Q instead of P_I and Q_I, to simplify the notation.

The cycle matrix is typically computed using a spanning tree. We now want to consider the effect of the formulas (44) and (45) if another spanning ditree would have been used. Clearly, this is about the change of the basis of the cycle space of the digraph G. So suppose we have two cycle

[31] The arcs in G indicate which nodes are compared, in measurements, periods, countries, etc. As a digraph it allows to distinguish (a, b) from (b, a), to indicate from which point a measurement was made (say to measure a height difference) or which is the base period and which is the comparison period (in price index applications). In other applications is may also be necessary to use the asymmetric role played by the nodes when they are compared directly.

matrices C_1 and C_2 produced for a digraph G, as a result of choosing two different spanning ditrees. In (31) it is shown how they are connected via a nonsingular matrix Ω: $C_2 = \Omega C_1$. Let Q_1 be the expression for (45) using $C = C_1$. Likewise we define Q_2 using $C = C_2$. Now we find, using (45)

$$\begin{aligned} Q_2 &= C_2'(C_2C_2')^{-1}C_2 = (\Omega C_1)'((\Omega C_1(\Omega C_1)')^{-1}\Omega C_1 \\ &= C_1'\Omega'(\Omega')^{-1}(C_1C_1')^{-1}\Omega^{-1}\Omega C_1 = C_1'(C_1C_1')^{-1}C_1 = Q_1. \end{aligned} \tag{46}$$

So (46) shows that Q in (45) is independent of the choice of a particular basis for the cycle space of G. Likewise P in (44), of course. Put differently, (44) and (45) are invariant under the transformation indicated in (31).

We now take a closer look at the matrix $\Gamma = CC'$ which appears in the cycle method, in particular in the matrices P in (44) and Q in (45). In particular we want to better understand the meaning of its entries. These can be divided into two groups: those on the main diagonal, and those not on the main diagonal. Γ_{ii} is the length of the i-th elementary cycle in the cycle matrix C, which is measured by counting the number of arcs it contains. Γ_{ij} is a measure for the overlap of the two elementary cycles i and j. The overlap may by 0 (no arcs in common), negative (with arc pairs in the respective elementary cycles, that is an arc and its counter-arc) or positive (arcs in common). Concisely expressed:

$$\Gamma_{ij} = \sigma(\bar{\gamma}_i, \bar{\gamma}_j)|\gamma_i \cap \gamma_j|, \tag{47}$$

where $\sigma(\bar{\gamma}_i, \bar{\gamma}_j)$ is the sign of the overlap of $\bar{\gamma}_i$ and $\bar{\gamma}_j$ taking the relative orientations of the arcs in the overlap into account, and where $|\gamma_i \cap \gamma_j|$ denotes the number of edges in common of the undirected cycles γ_i and γ_j, corresponding to $\bar{\gamma}_i$ and $\bar{\gamma}_j$, respectively. We have

$$\sigma(\bar{\gamma}_i, \bar{\gamma}_j) = \begin{cases} -1 & : \text{ when the overlap has opposite orientation,} \\ 1 & : \text{ when the overlap has the same orientation,} \end{cases} \tag{48}$$

where the overlap is considered for the undirected cycles γ_i and γ_j.[32)]

$|\Gamma_{ij}| = |\gamma_i \cap \gamma_j|$ is the number of common arcs, or pairs of arcs and counter-arcs, that is, edges. We have the following (crude) inequalities

$$0 \leq |\Gamma_{ij}| \leq \min\{|\Gamma_{ii}|, |\Gamma_{jj}|\}, \tag{49}$$

as the overlap of two elementary cycles cannot exceed the length of the smallest of the two.

In fact, we[33)] can sharpen the lower bound of $|\Gamma_{ij}|$ in (49):

32) It appears to be impossible that in an overlap of two elementary cycles in a digraph some arcs have the same orientation and others have the opposite orientation, but we have no proof of this statement.

33) As pointed out by Sander Scholtus.

$$\max\{0, |\Gamma_{ii}| + |\Gamma_{jj}| - m\} \leq |\Gamma_{ij}|. \tag{50}$$

where m is the number of edges in G. This follows from

$$m \geq |\gamma_i \cup \gamma_j| = |\gamma_i| + |\gamma_j| - |\gamma_i \cap \gamma_j| = |\Gamma_{ii}| + |\Gamma_{jj}| - |\Gamma_{ij}| \tag{51}$$

We now consider an example to illustrate how elementary cycles can intersect in a digraph.

Example Consider the digraph pictured in Figure 5.1. There are four arcs not on the spanning ditree: $(1,2)$, $(4,6)$, $(6,7)$ and $(6,5)$. Adding each of them yields an elementary cycle:

adding arc $(1,2)$: elementary cycle $(1,2,3) = \{(1,2),(2,3),(3,1)\}$
adding arc $(4,6)$: elementary cycle $(3,4,6) = \{(3,4),(4,6),(6,3)\}$
adding arc $(6,7)$: elementary cycle $(3,6,7) = \{(3,6),(6,7),(7,3)\}$
adding arc $(6,5)$: elementary cycle $(3,4,5,6) = \{(3,4),(4,5),(5,6),(6,3)\}$

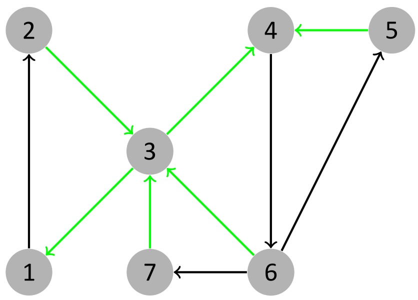

Figure 5.1 A graph with spanning tree (edges coloured green).

If we consider elementary cycles $(1,2,3)$ and $(3,4,6)$ we see that there is no overlap of arcs, irrespective of the orientation. If we consider elementary cycles $(3,4,6)$ and $(3,6,7)$ we see there is arc $(6,3)$ in the first elementary cycle and arc $(3,6)$, its counter-arc, in the second elementary cycle, which correspond to the common edge $\{3,6\}$. If we consider elementary cycles $(3,4,6)$ and $(3,4,5,6)$, we see that they have the arcs $(3,4)$ and $(6,3)$ in common. So this example has all three possible types of intersecting cycles. □

We now consider CC'. This is a symmetric matrix and hence diagonalizable. So there is an orthogonal matrix F_C such that

$$F_C' C C' F_C = \Lambda_C, \tag{52}$$

where Λ_C is a diagonal matrix. In fact, the columns of F_C are the eigenvectors of CC'. As $F_C F_C' = F_C' F_C = I$, so that $F_C^{-1} = F_C'$, it follows from (52) that

$$CC' = F_C \Lambda_C F_C'. \tag{53}$$

As CC' is positive definite, that is, $x'CC'x > 0$ for $x > 0$, we conclude that its eigenvalues λ are positive, i.e. $\lambda > 0$.

Now we consider Q as defined in (45)

$$\begin{aligned} x'Qx &= x'C'(CC')^{-1}Cx = x'C'(F_C\Lambda_C F_C')^{-1}Cx = x'C'F_C\Lambda_C^{-1}F_C'Cx \\ &= (F_C'Cx)'\Lambda_C^{-1}(F_C'Cx) = y'\Lambda_C^{-1}y \geq 0, \end{aligned} \tag{54}$$

where

$$y = F_C'Cx. \tag{55}$$

(54) shows that Q is positive semi-definite. But in fact Q is positive definite, that is, $x'Qx > 0$ if $x > 0$.

Example We consider the matrix C in (30) and compute the eigenvalues of CC':

$$CC' = \begin{pmatrix} 3 & 2 & 0 \\ 2 & 4 & -1 \\ 0 & -1 & 3 \end{pmatrix}. \tag{56}$$

We find[34)]

$$\Lambda_C = \begin{pmatrix} 5.791 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1.209 \end{pmatrix}, \tag{57}$$

where the entries on the main diagonal (that is, the eigenvalues) have been rounded to 3 decimal places, and the corresponding orthogonal matrix

$$F_C = \begin{pmatrix} 0.559 & 0.447 & -0.698 \\ 0.780 & 0.000 & 0.625 \\ -0.280 & 0.894 & 0.349 \end{pmatrix}, \tag{58}$$

where the entries have also been rounded to 3 decimal places. □

34) Using the eigen() function in the statistical package R, which computes both the eigenvalues and the eigenvectors of a matrix.

5.3 Cut method

The cut method is of more recent origin than the cycle method. In fact it was developed when the present paper was written. Both methods are linked.

The cycle method was originally intended to correct for measurement errors in height difference measurements. 'Height' plays the role of potential. The cut method can be seen as a method to correct a kind of flow in a network, where there is inflow and outflow at each node. This may not be in balance. Suppose that our v-digraph represents a street network, with traffic flow in each street as the 'valuation'. Suppose furthermore that these traffic flows per street have been measured independently of each other. At each node we would like to have that the total inflowing traffic equals the total outflowing traffic. As we are dealing with measurements this may not be the case with the measured data, at all nodes. So there is an incentive to mend such inconsistencies and demand that Kirchhoff's law ('inflow = outflow') holds at each node. Using linear regression analysis in a similar way as for the cycle method, we can formulate a constrained linear regression method, with linear constraints, this time dictated by the inflow and outflow requirements at each node. These constraints can be expressed using a matrix D instead of the matrix C. D is obtained from the incidence matrix $\mathcal{I}$ by deleting a single row, for instance the last one. If we assume the weight matrix to be equal to the identity matrix, we can write down the solution as

$$\check{x} = x - D'(DD')^{-1}Dx = (I - D'(DD')^{-1}D)x = Rx, \tag{59}$$

by looking at the result for the cycle method and by simply replacing C by D, where

$$R = I - D'(DD')^{-1}D. \tag{60}$$

In analogy to Q we define

$$S = D'(DD')^{-1}D. \tag{61}$$

Now the adjusted values $\check{x}$ as defined in (59) satisfy

$$D\check{x} = 0, \tag{62}$$

which is a formal 'translation' of Kirchhoff's law for each node in G.

As in case of the cycle method (see (46)), we can show that the choice of basis of the cut space of G does not affect D: the expression $D'(DD')^{-1}D$ remains invariant.

The matrix $\Theta = DD'$ is related to the graph Laplacian associated with the underlying graph of the v-digraph involved. The row space of D is equal to that of its incidence matrix $\mathcal{I}$. The difference is that the rows in D are independent and hence form a basis of that space. In case of $\mathcal{J}$ the row vectors are dependent. It follows that the eigenvalues of Θ are eigenvalues of $\mathcal{J}\mathcal{J}'$ as well, which is the graph Laplacian (see (17)). These eigenvalues are positive.[35)]

We now have a closer look at Θ, which is the counter-part of $\Gamma = CC'$ (see the end of Section 5.2). Again as with CC' we make a distinction between the entries on the main diagonal, Θ_{ii} and the off-diagional entries Θ_{ij} with $i \neq j$. Θ_{ii} is the number of ingoing arcs plus the number of outgoing arcs of node i, which is the same as the degree of node i of the underlying graph of G. In case of Θ_{ij} we make a distinction between adjacent nodes i and j and non-adjacent ones. In the first case $\Theta_{ij} = -1$ as there is arc (i,j) or its counter-arc (j,i), with opposite orientation. Arc (i,j) can be viewed as an arc leaving i and entering j. For arc (j,i) the opposite is the case. So in both cases we have different signs, '$+$' and '$-$', resulting in entries $\Theta_{ij} = -1$ and $\Theta_{ji} = -1$. In case i and j are non-adjacent, $\Theta_{ij} = 0$, as there is no arc starting in i and ending in j, or vice versa. Concisely

$$\Theta_{ij} = \begin{cases} \text{degree}(i) & : \quad i = j \\ -1 & : \quad (i,j) \in E \text{ or } (j,i) \in E \\ 0 & : \quad (i,j) \notin E \text{ and } (j,i) \notin E \end{cases} \tag{63}$$

We now consider an example that can be compared to the one at the end of Section 5.2, but this time with the focus on D instead of C.

Example We consider the matrix D in (32) and compute the eigenvalues of

$$DD' = \begin{pmatrix} 2 & -1 & -1 & 0 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 4 & -1 \\ 0 & -1 & -1 & 3 \end{pmatrix}. \tag{64}$$

We find

$$\Lambda_D = \begin{pmatrix} 4.890 & 0 & 0 & 0 \\ 0 & 4.208 & 0 & 0 \\ 0 & 0 & 2.493 & 0 \\ 0 & 0 & 0 & 0.409 \end{pmatrix}, \tag{65}$$

where the entries have been rounded to 3 decimal places.

The corresponding orthogonal matrix is

35) As $\mathcal{J}$ is not of full row rank, so $\mathcal{J}\mathcal{J}'$ is not of full rank, implying that the kernel of $\mathcal{J}\mathcal{J}'$, i.e. Ker $\mathcal{J}\mathcal{J}'$, is a linear subspace of dimension at least equal to 1.

$$F_D = \begin{pmatrix} 0.264 & 0.268 & -0.695 & -0.613 \\ 0.107 & -0.802 & 0.214 & -0.547 \\ -0.870 & 0.210 & 0.129 & -0.428 \\ 0.403 & 0.490 & 0.675 & -0.376 \end{pmatrix}, \quad (66)$$

where the entries have also been rounded to 3 decimal places. □

5.4 Relationship between cycle method and cut method

5.4.1 Identity matrix as weight matrix

The cycle methode and the cut method are intimately related. We have the following key matrices associated with these methods, introduced earlier in the text (see (44), (45), (60) and (61):

$$P = I - C'(CC')^{-1}C$$
$$Q = C'(CC')^{-1}C.$$
$$R = I - D'(DD')^{-1}D,$$
$$S = D'(DD')^{-1}D.$$

Identity (33) is an important one connecting the cycle method and the cut method. Using the identities (44), (45), (60), (61) and (33) we can derive others, several of which have been collected in Table 5.1. Other algebraic properties of these matrices have been added as well. Note that all matrices involve only simple rational operations to compute them from the matrices C or D.

$P^2 = P$	$Q^2 = Q$	$DR = 0$	$RD' = 0$
$R^2 = R$	$S^2 = S$	$DS = D$	$SD' = D'$
$P + Q = I$	$R + S = I$	$CS = 0$	$SC' = 0$
$PQ = 0$	$QP = 0$	$DQ = 0$	$QD' = 0$
$RS = 0$	$SR = 0$	$DC' = 0$	$CD' = 0$
$P' = P$	$Q' = Q$	$QS = 0$	$SQ = 0$
$R' = R$	$S' = S$	$PS = S$	$SP = S$
$CP = 0$	$PC' = 0$	$RQ = Q$	$QR = Q$
$CQ = C$	$QC' = C'$	$PR = R - Q$	$RP = P - S$
$DP = D$	$PD' = D'$	$CR = C$	$RC' = C'$
$P = S$	$Q = R$		

Table 5.1 Algebraic identities for matrices related to the cycle method and the cut method.

The identities $P = S$ and the equivalent $Q = R$ at the bottom line of Table 5.1 result from observations about the cycle space and cut space of a digraph G in Section 4.3. These identities plus the identities $DC' = 0$ and the equivalent $CD' = 0$ require insight into the construction and meaning of the cycle space and cut space of a digraph. The remaining identities in this table follow from simple algebraic computations.

No attempt was made in Table 5.1 to list all possible identities of the matrices involved in the definition of the cycle method and the cut method. Nor was it attempted to look for a minimum set of identities, from which the remaining ones can be derived.

Now $P = S$ (or $Q = R$) implies the remarkable identity:

$$I = C'(CC')^{-1}C + D'(DD')^{-1}D. \tag{67}$$

Identity (67) can in turn be used to derive the related equalities $DC' = 0$ (see (33)) and $CD' = 0$.

6 Helmholtz decomposition for v-digraphs

In Section 2 the Helmholtz decomposition of a vector field was shown. It consisted of two components, one that was rotation free and one that had zero divergence. In the present section we explore such a decomposition for v-digraphs using the cycle method and the cut method, which actually allow one to compute the two components.

In this case we consider any digraph with arc valuation $G = (V, E, x)$, with $x : E \to \mathbb{R}$ a valuation associated with the arcs of G. We consider a transform of x, inspired by the cycle method:

$$\hat{x} = Px = (I - C'(CC')^{-1}C)x, \tag{68}$$

We do not use a weight matrix W as in the cycle method. Or, more precisely, we assume it equals I, the identity matrix (of the appropriate order, depending on the application at hand).

Now (68) leads to a decomposition of the valuation x into two components:

$$x = \left(x - C'(CC')^{-1}Cx\right) + C'(CC')^{-1}Cx = Px + Qx = \hat{x} + Qx, \tag{69}$$

Our aim is to show that decomposition (69) is the equivalent of the Helmholtz decomposition for a vector field, but now applied to a v-digraph: Px is the component derived from a potential function defined on the nodes of the digraph, whereas Qx is the 'divergence free' component, which means that it has no sources ('only values out') or sinks ('only values in') at any node of the digraph. Phrased differently, it means that for each node the total of the values into a node equal the total of the values out of the node (Kirchhoff's law from electrical circuit theory).

It follows that for the component Px holds

$$C\hat{x} = CPx = 0, \tag{70}$$

which states that for any elementary cycle the sum of the values associated with its arcs is zero. This implies that this 'zero sum' property actually holds for any cycle in G.

Remark In vector analysis (in particular in applications such as electricity or gravitation) this means that the integral of a closed path at any point in the field is zero.[36)] Characteristic for a conservative field is that it implies the existence of a potential function from which the field (φ) can be derived by taking the gradient (in fact, $-\nabla\varphi$). The reverse is also true: if a field is derived from a potential, it is conservative. In the digraph case, Px can be compared to the field derived from a potential function defined on the nodes of G (cf. [12]). ☐

We now consider the second component in the final expression in (69), i.e. Qx. We want to show that the v-digraph equivalent of the divergence of Qx is zero, which means that inflow equals outflow at each node (i.e. Kirchhoff's law). In fact, because $DC' = 0$ (see identity (33), it follows that

$$DQx = 0, \tag{71}$$

which proves our contention.

With (70) and (71) we have obtained the defining properties of the components Px and Qx: Px is the component which can be derived from a potential function defined on the nodes, and Qx is the divergence free component, for which inflow = outflow at each node. So (69) is the Helmholtz decomposition for v-digraphs that we were looking for. And the cycle method produces it.

But so does the cut method. Using (59) we have the following decomposition

$$x = \left(I - D'(DD')^{-1}D\right)x + D'(DD')^{-1}Dx = Rx + Sx, \tag{72}$$

where R and S are as defined in (60) and (61).

Because $P = S$ and $Q = R$ (cf. Table 5.1) the decomposition (72) is also a Helmholtz decomposition, this entirely defined in terms of D instead of C.

We now give an example of a Helmholtz decomposition to illustrate this identity by direct computation.

Example We consider the matrices C and D as given in (30) and (32), which are both associated with the digraph in Figure 3.1. We then find

36) Such a field is called conservative. The gravitational field and the electric field are well-known examples of conservative fields.

$$Q = C'(CC')^{-1}C$$

$$\approx \begin{pmatrix} 0.381 & -0.381 & -0.048 & -0.095 & -0.238 & 0.143 & -0.048 \\ -0.381 & 0.381 & 0.048 & 0.095 & 0.238 & -0.143 & 0.048 \\ -0.048 & 0.048 & 0.381 & -0.238 & -0.095 & -0.143 & 0.381 \\ -0.095 & 0.095 & -0.238 & 0.524 & -0.19 & -0.286 & -0.238 \\ -0.238 & 0.238 & -0.095 & -0.19 & 0.524 & 0.286 & -0.095 \\ 0.143 & -0.143 & -0.143 & -0.286 & 0.286 & 0.429 & -0.143 \\ -0.048 & 0.048 & 0.381 & -0.238 & -0.095 & -0.143 & 0.381 \end{pmatrix}. \tag{73}$$

and

$$S = D'(DD')^{-1}D$$

$$\approx \begin{pmatrix} 0.619 & 0.381 & 0.048 & 0.095 & 0.238 & -0.143 & 0.048 \\ 0.381 & 0.619 & -0.048 & -.095 & -0.238 & 0.143 & -0.048 \\ 0.048 & -0.048 & 0.619 & 0.238 & 0.095 & 0.143 & -0.381 \\ 0.095 & -0.095 & 0.238 & 0.476 & 0.19 & 0.286 & 0.238 \\ 0.238 & -0.238 & 0.095 & 0.19 & 0.476 & -0.286 & 0.095 \\ -0.143 & 0.143 & 0.143 & 0.286 & -0.286 & 0.571 & 0.143 \\ 0.048 & -0.048 & -0.381 & 0.238 & 0.095 & 0.143 & 0.619 \end{pmatrix}, \tag{74}$$

where all entries have been rounded to 3 decimal places.

Indeed, $Q + S = I$ (apart from rounding errors), as (67) requires.[37)] For a valuation x on the arcs of G the components of the Helmholtz decomposition are: Qx ('divergence $= 0$ component') and Sx ('component derived from a potential'). □

Remark Identity (67) can be used to determine the cycle estimator. We can use D obtained from the incidence matrix $\mathcal{I}$ by extracting a basis for the row space of $\mathcal{I}$. This avoids calculating a basis for the cycle space (and hence computing C) using a spanning tree. Unfortunately this 'trick' can only be used in case $W = I$, as is the case with the Helmholtz decomposition. □

This concludes the present paper: we have shown in this paper how to translate the Helmholtz decomposition in a vector field to a similar construct on a v-digraph. We also have shown that the decomposition exists for such finite structures and that it is unique for a v-digraph, how it depends on its topology (cycle structure) and how it can be computed using the cycle method or the cut method. The cycle method was known long before the present paper was written (in a very different context). The cut method was developed as part of the work on the present paper.

7 Discussion

In this paper it was shown how the Helmholtz decomposition for vector fields can be translated to v-digraphs in a natural way. Also it was shown how the cycle method and the cut method can

37) If the numbers in (73) and (74) had been stated in greater precision, rounding errors would probably have appeared, although they would have been very small.

be used to compute the decomposition in the graph theoretical setting. The cycle method was developed by the author many years ago in a completely different setting and was applied by him in another area of statistics, namely price index numbers. The cut method was developed in the present paper as a complement to the cycle method.

In the present application the cycle method is applied in a specific way, with the weight matrix $W = I$, where I the identity matrix of the same order as W. In fact this is necessary for the required decomposition to hold. It follows that the Helmholtz decomposition for digraphs only depends on the cycle matrix C and the cut matrix D. These are (topological) properties of the v-digraph used. In the general cycle method the weight matrix is a choice of the analyst applying the method.

If a tool would be available that is an implementation of the cycle method and/or the cut-method then this tool could be used to compute the Helmholtz decomposition in v-digraphs (among other things). For the present paper computational issues are outside its scope. However, development of such a tool is highly desirable. It makes live much easier for the users, as its implementation requires knowledge that is somewhat specialized and not part of the skill set of the average statistician. Also, such a tool would be useful for price index applications (to transitivize nontransitive price index figures.

References

[1] B. Balk (2008). *Price and Quantity Index Numbers*. Cambridge University Press.

[2] B. Biyikoğlu, J. Leydold & P. Stadler (2007). *Laplacian Eigenvectors of Graphs – Perron-Frobenius and Faber-Krahn Type Theorems*, Lecture Notes in Mathematics, Vol. 1915. Springer.

[3] A. Gibbons (1985). *Algorithmic Graph Theory*. Cambridge University Press.

[4] C. Godsil & G. Royle (2001). *Algebraic Graph Theory*. Springer.

[5] F. Graybill (1983). *Matrices with Applications in Statistics*, 2^{nd} ed. Duxbury.

[6] H. Helmholtz (1858). Über Integrale der Hydrodynamischen Gleichungen, welcher der Wirbelbewegungen Entsprechen, *Journal für de Reine und Angewandte Mathematik*, 55, pp. 25-55.

[7] W.V.D. Hodge (1952). *The Theory and Applications of Harmonic Integrals*. Cambridge University Press.

[8] L.-H. Lim (2010). Rank Aggregation via Hodge Theory, Slides, August 18, University of Chicago.

[9] L.-H. Lim (2020). Hodge Laplacians on Graphs, *Siam Review*, 62, pp. 685-715.

[10] A. Strang (2020). Applications of the Helmholtz-Hodge Decomposition to Networks and Random Processes, PhD dissertation, Division of Mathematics, Applied Mathematics and Statistics, Case Western Reserve University.

[11] L. Willenborg (1993). An Adjustment Method Based on Graph Homology, Report, CBS, Voorburg.

[12] L. Willenborg (2011). Chain Indices and Path Independence, Discussion paper, CBS, The Hague.

[13] L. Willenborg (2017a). Transitivizing Elementary Price Indices for Internet Data Using the Cycle Method, Discussion paper, CBS, The Hague.

[14] L. Willenborg (2017b). From GEKS to Cycle Method, Discussion paper, CBS, The Hague.

[15] L. Willenborg (2022). Elastic Energy of Networks, Discussion paper, CBS, The Hague.

Colophon

Publisher
Statistics Netherlands
Henri Faasdreef 312, 2492 JP The Hague
www.cbs.nl

Prepress
Statistics Netherlands, Grafimedia

Design
Edenspiekermann

Information
Telephone +31 88 570 70 70, fax +31 70 337 59 94
Via contact form: www.cbs.nl/information