A Prime Factor Theorem for a Generalized Direct Product

Wilfried Imrich[a] and Peter F. Stadler[b,c]

[a]Department of Mathematics and Information Technology
Montanuniversität Leoben,
Franz Josef-Straße 18, A-8700 Leoben, Austria
Tel: +43-3842-402-1700, Fax: +43-3842-402-1702
E-Mail: imrich@unileoben.ac.at

[b]Bioinformatics Group, Department of Computer Science
University of Leipzig,
Härtelstrasse 16-18, D-04107 Leipzig, Germany
Tel: +49-341-97-16691, Fax: +49-341-97–16709,
URL: http://www.bioinf.uni-leipzig.de/~studla

[c]The Santa Fe Institute,
1399 Hyde Park Road, Santa Fe, NM 87501, USA

Abstract

We introduce the concept of neighborhood systems as a generalization of directed, reflexive graphs and show that the prime factorization of neighborhood systems with respect to the the direct product is unique under the condition that they satisfy an appropriate notion of thinness.

Keywords: products, set systems, prime factor theorem
AMS Subject Classification (2000): 05C20, 05C65, 05C70

1. Introduction

In this contribution we consider a special class of set systems that arises in a natural way in theoretical biology. Many central notions in evolutionary biology are intrinsically topological: for example, one speaks of "continuous" and "discontinuous" transitions in evolution of phenotypes (i.e., the organization and physical shape of an organism). Since genetic variation is determined by mutation, recombination, and other genetic operators acting on the sequence of the organism's genotype (i.e., its DNA sequence) it becomes natural to organize the phenotypes according to their genetic accessibility [2]. Computer models shows that the resulting finite spaces lack e.g. a metric structure. Instead, they are most conveniently formalized by means of generalized closure spaces, that is, structures that generalize topological spaces, see

e.g. [3, 1]. To this end one defines the closure $c(A)$ of a set of organisms A as the set of organisms that can be obtained from A by application of the genetic operators [7, 6].

In the simplest case, where only mutations are considered at the phenotypic level, one has to deal with pretopological spaces; here the closure function is not assumed to be idempotent, and (equivalently) a neighborhood of a point does not necessarily contain an open neighborhood [1]. A further generalization is necessary to incorporate recombination. The resulting *neighborhood spaces* are defined by neighborhood systems that are no longer proper filters but are merely arbitrary isotonic set systems.

In [8] a theory is discussed in which the notion of a "biological character" is identified with a factor of the phenotype spaces with respect to a suitable generalization of the topological product, which reduces to direct product of directed graphs with loops (i.e., finite pretopological spaces) in the mutation-only case. The crucial observation for the biological interpretation is the existence of a unique prime factor decomposition and a unique coordinatization of the graph under certain circumstances [5].

In this contribution we consider the more general case of finite neighborhood spaces. The corresponding generalization of directed graphs, for which we propose the term $\mathfrak{N}-$*systems*, has, to our knowledge, not been investigated so far.

2. Definitions and Preliminaries

Definition 1. *An* $\mathfrak{N}$-system *consists of a nonempty finite set* X *and a system* $\mathfrak{N}$ *of collections of subsets of* X *that associates to each* $x \in X$ *a collection* $\mathfrak{N}(x) = \{N^1(x), N^2(x), \ldots, N^{d(x)}\}$ *of* $d(x) \geq 1$ *subsets of* X *with the following properties:*

(N0) $\mathfrak{N}(x) \neq \emptyset$.
(N1) $N^i(x) \subseteq N^j(x)$ *implies* $i = j$.
(N2) $x \in N^i(x)$ *for* $1 \leq i \leq d(x)$.

We also call an $\mathfrak{N}$*-system* neighborhood system *and denote it by* $(X, \mathfrak{N})$.

Remark. If $d(x) = 1$ for all $x \in X$ then $(X, \mathfrak{N})$ describes a directed graph with loops. In this case $\mathfrak{N}(x) = \{N^1(x)\}$ for every x and $N^1(x)$ is the neighborhood of x.

Definition 2. *Let* $(X_1, \mathfrak{N}_1)$ *and* $(X_2, \mathfrak{N}_2)$ *be two* $\mathfrak{N}$*-systems. We define their direct product* $(X_1, \mathfrak{N}_1) \times (X_2, \mathfrak{N}_2)$ *in the following way:*

(1) *The vertex set is* $X = X_1 \times X_2$.
(2) *The neighborhoods* $\mathfrak{N}(x_1, x_2)$ *are the sets* $\{N' \times N'' | N' \in \mathfrak{N}_1(x_1),\ N'' \in \mathfrak{N}_2(x_2)\}$.

Lemma 3. *The direct product of two* $\mathfrak{N}$-systems *is an* $\mathfrak{N}$-system.

Proof. Clear. □

If $\mathfrak{N}_1(x_1) = \{N(x_1)\}$ and $\mathfrak{N}_2(x_2) = \{N(x_2)\}$ then $\mathfrak{N}(x_1, x_2) = \{N(x_1) \times N(x_2)\}$. Hence, if $(X_1, \mathfrak{N}_1)$ and $(X_2, \mathfrak{N}_2)$ both represent graphs, then their product also represents a graph, which is the direct product of graphs in the usual sense.

Definition 4. *Let* $\Gamma(X, \mathfrak{N})$ *be the directed graph (with loops) with vertex set* X *and edge set*

$$E = \left\{ (x, y) \,\middle|\, y \in \bigcup_{i=1}^{d(x)} N^i(x) \right\} \tag{1}$$

We say that $(X, \mathfrak{N})$ *is connected if* $\Gamma(X, \mathfrak{N})$ *is connected.*

Lemma 5. $\Gamma((X_1, \mathfrak{N}_1) \times (X_1, \mathfrak{N}_1)) = \Gamma(X_1, \mathfrak{N}_1) \times \Gamma(X_2, \mathfrak{N}_2)$.

Proof. The vertex set of both graphs is $X_1 \times X_2$. In order to see that the edge sets are the same it suffices to observe that

$$\left(\bigcup_i N_i(x) \right) \times \left(\bigcup_j N_j(y) \right) = \bigcup_i \bigcup_j (N_i(x) \times N_j(y)) \tag{2}$$

is true in general. □

Let Γ be a directed graph (X, E). We say two vertices u, v of Γ are *equivalent*, in symbols uRv, if $(u, x) \in E \Leftrightarrow (v, x) \in E$ and $(x, u) \in E \Leftrightarrow (x, v) \in E$. The graph Γ is *thin* if R is the identity relation on X.

It is clear what we mean by Γ/R and that Γ/R is always thin. For us the following relation is of particular importance

$$(X \times Y)/R = X/R \times Y/R.$$

For thin connected reflexive relations (i.e., thin digraphs with loops at each vertex) we have the strong refinement property (McKenzie [5]), which guarantees a unique coordinatization with respect to the direct product. In fact, McKenzie's result guarantees the existence of a common refinement for any two decompositions with respect to the direct product, but this of course implies unique prime factorization for finite structures.

3. The main result

Theorem 6. *If the* $\mathfrak{N}$*-system* $(X, \mathfrak{N})$ *has a thin, connected digraph* $\Gamma(X, \mathfrak{N})$. *Then* $(X, \mathfrak{N})$ *has a unique prime factorization with unique coordinatization.*

Proof. McKenzie's result guarantees a unique decomposition of $\Gamma(X, \mathfrak{N})$ with unique coordinatization. If $(X_1, \mathfrak{N}_1)$ is a factor of $(X, \mathfrak{N})$, then $\Gamma(X_1, \mathfrak{N}_1)$ is a factor of $\Gamma(X, \mathfrak{N})$. However, it is possible that $(X, \mathfrak{N})$ or any of its $\mathfrak{N}$-factors $(X_i, \mathfrak{N}_i)$ is indecomposable (as an $\mathfrak{N}$-system) but that $\Gamma(X, \mathfrak{N})$ or $\Gamma(X_1, \mathfrak{N}_1)$ can be further decomposed as a graph. It follows that the decomposition of $(X, \mathfrak{N})$ will in general be coarser than the decomposition of $\Gamma(X, \mathfrak{N})$.

Since the coordinatization is unique (up to the order of the factors), the projections $N_i^k(x_i)$ of $N^k(x)$ to the prime factors Γ_i are unique. This implies that an $\mathfrak{N}$-system $(X, \mathfrak{N})$ will be prime in the case of a decomposable $\Gamma(X, \mathfrak{N})$ if there is an $N^k(x)$ that is not the product $\prod_i N_i^k(x_i)$ of its projections into the factors Γ_i. In this case $N^k(x)$ is a proper subset of $\prod_i N_i^k(x_i)$.

Since the decomposition of $(X, \mathfrak{N})$ is coarser in general we need to show that there is a unique finest decomposition of $(X, \mathfrak{N})$. For this it suffices to show that there is a common $\mathfrak{N}$-refinement to any two $\mathfrak{N}$-system decompositions of $(X, \mathfrak{N})$.

Suppose $X = Y \times Z$ and $X = U \times V$ are two such decompositions. Clearly there is a Γ-refinement with factors Y_U, Y_V, Z_U, and Z_V, such that $Y = Y_U \times Y_V$, $Z = Z_U \times Z_V$, $U = Y_U \times Z_U$, and $V = Y_V \times Z_V$. Given a vertex $v \in X$ we call the subgraph of X that is induced by the set of vertices that differ from v only in the Y-coordinate the *Y-layer* of through v. Clearly every Y-layer Y^v is the product of its projections into Y_U and Y_V,

$$Y^v = p_{Y_U}(Y^v) \times p_{Y_V}(Y^v).$$

Let A be an arbitrary $N^i(x)$. By assumption

$$A = p_Y A \times p_Z A = p_U A \times p_V A.$$

Clearly $p_Y A$ is a subset of $p_U p_Y A \times p_V p_Y A$. We have to show that equality holds. It is equivalent to the statement that every Y-layer of A is the product of its projections into U and V. In other words, every Y-layer of A is a subproduct of A with respect to the decomposition $A = p_U A \times p_V A$.

Suppose this is not the case for the Y-layer $(p_Y A)^v$ of A. Then there exists a vertex $w = (w_U, w_V) \notin (p_Y A)^v$ and vertices $a, b \in (p_Y A)^v$, where $a_U = w_U$ and $b_V = w_V$. Note that $w \notin Y^v$ since $p_Y A^v = Y^v \cap A$. Furthermore a and b have the same Z-coordinate. Thus,

$$a = (w_U, a_V) = (w_{Z_U}, w_{Y_U}, a_{Z_V}, a_{Y_V}), \qquad b = (b_U, w_V) = (b_{Z_U}, b_{Y_U}, w_{Z_V}, w_{Y_V}).$$

Their Z-coordinates are $a_Z = (w_{Z_U}, a_{Z_V})$ and $b_Z = (b_{Z_U}, w_{Z_V})$. Since they are equal, they are equal to (w_{Z_U}, w_{Z_V}), which is the Z-coordinate (w_{Z_U}, w_{Z_V}) of w. This is not possible because $w \notin (p_Y A)^v$. Hence, the Γ-refinement is also an $\mathfrak{N}$-refinement and $\mathfrak{N}$-prime factorization is unique for thin structures. □

We call two vertices u, v of a neighborhood system equivalent, in symbols uRv, if $\mathfrak{N}(u) = \mathfrak{N}(v)$ and if $u \in N^i(z) \Leftrightarrow v \in N^i(z)$ for all $z \in X$ and all $i \in \{1, \ldots, d(z)\}$. $(X, \mathfrak{N})$ is *thin* if R is the identity relation.

Note that $(X, \mathfrak{N})$ can be thin, even when $\mathfrak{N}(u) = \mathfrak{N}(v)$ for certain pairs of vertices. In such a case there must exist an $N^i(z)$ that contains exactly one of the vertices u, v.

Conjecture 7. *If $(X, \mathfrak{N})$ is connected, then it has a unique prime factor decomposition with respect to the direct product of $\mathfrak{N}$-systems. If $(X, \mathfrak{N})$ is thin and connected, then we also have unique coordinatization.*

With methods from [4] we can show that the conjecture holds if $\mathfrak{N}(u) = \mathfrak{N}(v)$ implies $u = v$, but the proof is long and tedious.

Acknowledgments. This work was supported in part by the Bioinformatics Initiative of the *Deutsche Forschungsgemeinschaft*, grant no. BIZ-6/1-2.

References

[1] E. Čech. *Topological Spaces*. Wiley, London, 1966.
[2] W. Fontana and P. Schuster. Continuity in Evolution: On the Nature of Transitions. *Science*, 280:1451–1455, 1998.
[3] P. C. Hammer. Extended topology: Continuity I. *Portug. Math.*, 25:77–93, 1964.
[4] W. Imrich and S. Klavžar. *Product graphs*. Wiley-Interscience Series in Discrete Mathematics and Optimization. Wiley-Interscience, New York, 2000. Structure and recognition, With a foreword by Peter Winkler.
[5] R. McKenzie. Cardinal multiplication of structures with a reflexive relation. *Fund. Math.*, 70:59–101, 1971.
[6] B. M. R. Stadler and P. F. Stadler. Generalized topological spaces in evolutionary theory and combinatorial chemistry. *J. Chem. Inf. Comput. Sci.*, 42:577–585, 2002.
[7] B. M. R. Stadler, P. F. Stadler, G. Wagner, and W. Fontana. The topology of the possible: Formal spaces underlying patterns of evolutionary change. *J. Theor. Biol.*, 213:241–274, 2001.
[8] G. Wagner and P. F. Stadler. Quasi-independence, homology and the unity of type: A topological theory of characters. *J. Theor. Biol.*, 220:505–527, 2003.