



IntechOpen

Topology
Recent Advances and Applications

Edited by Paul Bracken



Topology - Recent Advances and Applications

Edited by Paul Bracken

Published in London, United Kingdom

Topology – Recent Advances and Applications
<http://dx.doi.org/10.5772/intechopen.1000228>
Edited by Paul Bracken

Contributors

Naimat Ullah, Mohammed Shehu Shagari, Tahir Ahmad Khan, Aziz Ullah Khan, Muhammad Atta Ullah Khan, Mohammad S. R. Chowdhury, Muhammad Suhail Aslam, Musawa Yahya Almusawa, Muhammad Imran Asjad, Sergey Galaev, Evgeny Kokin, Emmanoel Ferreira, Sandra Nadim Aldarrouj, Ibrahim Alshannour, Xiaoquan Xu, Yousef Methkal Abd Algani, Amal Sharif-Rasslan, Paul Bracken, Paula Mellado, Roberto E. Troncoso, Edward J. Haug, Petar Pavesic, Kornél Kovách, Daniella Éva Pigniczki

© The Editor(s) and the Author(s) 2023

The rights of the editor(s) and the author(s) have been asserted in accordance with the Copyright, Designs and Patents Act 1988. All rights to the book as a whole are reserved by INTECHOPEN LIMITED. The book as a whole (compilation) cannot be reproduced, distributed or used for commercial or non-commercial purposes without INTECHOPEN LIMITED's written permission. Enquiries concerning the use of the book should be directed to INTECHOPEN LIMITED rights and permissions department (permissions@intechopen.com).

Violations are liable to prosecution under the governing Copyright Law.



Individual chapters of this publication are distributed under the terms of the Creative Commons Attribution 3.0 Unported License which permits commercial use, distribution and reproduction of the individual chapters, provided the original author(s) and source publication are appropriately acknowledged. If so indicated, certain images may not be included under the Creative Commons license. In such cases users will need to obtain permission from the license holder to reproduce the material. More details and guidelines concerning content reuse and adaptation can be found at <http://www.intechopen.com/copyright-policy.html>.

Notice

Statements and opinions expressed in the chapters are those of the individual contributors and not necessarily those of the editors or publisher. No responsibility is accepted for the accuracy of information contained in the published chapters. The publisher assumes no responsibility for any damage or injury to persons or property arising out of the use of any materials, instructions, methods or ideas contained in the book.

First published in London, United Kingdom, 2023 by IntechOpen
IntechOpen is the global imprint of INTECHOPEN LIMITED, registered in England and Wales, registration number: 11086078, 5 Princes Gate Court, London, SW7 2QJ, United Kingdom

British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library

Additional hard and PDF copies can be obtained from orders@intechopen.com

Topology – Recent Advances and Applications
Edited by Paul Bracken
p. cm.
Print ISBN 978-1-83769-559-1
Online ISBN 978-1-83769-558-4
eBook (PDF) ISBN 978-1-83769-560-7

We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,500+

Open access books available

176,000+

International authors and editors

190M+

Downloads

156

Countries delivered to

Our authors are among the
Top 1%
most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com



Meet the editor



Paul Bracken is a professor in the Department of Mathematics at the University of Texas Rio Grande Valley, Edinburg, USA. He received a BSc from the University of Toronto and a Ph.D. from the University of Waterloo, Canada. His research interests include applications of differential geometry and partial differential equations to problems in mathematical physics, as well as problems in quantum mechanics and quantum field theory. He has published well over 180 research papers in journals and books and presented talks at numerous meetings and conferences. He is also a member of many societies, including the AMS, CMS and APS. This is the ninth volume he has contributed in his association with IntechOpen.

Contents

Preface	XI
Section 1	
Topology and Fixed-Point Theorems	1
Chapter 1	3
Common Fixed Point Theorems in Complex-Valued Non-Negative Extended b-Metric Spaces <i>by Naimat Ullah, Mohammed Shehu Shagari, Tahir Ahmad Khan, Aziz Ullah Khan and Muhammad Atta Ullah Khan</i>	
Chapter 2	17
Fixed Point Theorems in Complex-Valued Double Controlled Metric-Like Spaces <i>by Mohammad S.R. Chowdhury, Muhammad Suhail Aslam, Musawa Yahya Almusawa and Muhammad Imran Asjad</i>	
Section 2	
Geometry and Topology	37
Chapter 3	39
Geometry of Sub-Riemannian Manifolds Equipped with a Quasi-Semi-Weyl Structure <i>by Sergey Galaev and Evgeny Kokin</i>	
Chapter 4	51
Obtaining the Main Curvatures of Orientable Hypersurfaces, Based on Differential Geometry <i>by Emmanoel Ferreira</i>	
Section 3	
Modern Research Topics in Topology	67
Chapter 5	69
Topology Optimization Generating System <i>by Sandra Nadim Aldarrouj and Ibrahim Alshannour</i>	

Chapter 6	87
Some Recent Advances in Non-Hausdorff Topology <i>by Xiaoquan Xu</i>	
Chapter 7	113
The Topological Connectivity of the Bipartite Graph's Independence Complex <i>by Yousef Methkal Abd Algani and Amal Sharif-Rasslan</i>	
Section 4	
Applications of Topology in Physics	125
Chapter 8	127
Anomalies and Their Occurrence in the Study of Topology in Yang-Mills Gauge Theories <i>by Paul Bracken</i>	
Chapter 9	143
Topological Phenomena in Spin Systems: Textures and Waves <i>by Paula Mellado and Roberto E. Troncoso</i>	
Chapter 10	167
Mechanical System Kinematics and Dynamics on Differentiable Manifolds <i>by Edward J. Haug and Petar Pavesic</i>	
Chapter 11	191
Advanced Knotting Techniques (Examples from Surgical Practice) <i>by Kornél Kovách and Daniella Éva Pigniczki</i>	

Preface

Topology largely deals with properties of configurations or geometric objects, such as point sets, that are unaltered by homeomorphisms. Topology is not only an independent area of study, but it also plays a fundamental role in other areas of mathematics such as analysis differential geometry and graph theory. The contributed papers in this book cover current topics in topology as well as its applications to specific areas of science such as mathematical physics and even biology. Among the contributions are papers concerning the impact of topology on fixed point theorems, advances in non-Hausdorff topology and topology optimization. Also included are topics directly overlapping with modern physics, such as topological phenomena in spin systems, and several originating in differential geometry.

This book is written by an international group of invited researchers and authors. It is a pleasure to thank them for their hard work and significant contributions to the present volume. I am grateful to IntechOpen for the opportunity to edit this volume on the subject of topology, and to publishing manager Mr. Josip Knapic for his assistance and help throughout the process.

Paul Bracken

Professor,

Department of Mathematics,

The University of Texas Rio Grande Valley,

Edinburg, Texas, USA

Section 1

Topology and Fixed-Point Theorems

Common Fixed Point Theorems in Complex-Valued Non-Negative Extended b-Metric Spaces

Naimat Ullah, Mohammed Shehu Shagari, Tahir Ahmad Khan, Aziz Ullah Khan and Muhammad Atta Ullah Khan

Abstract

We define modified F.P results for mappings, according to some rational contractions and present to C.C.N.V.b-M Spaces. The relevant F.P theorems in the context of extended b-metric b-metric and classical M.Spaces are improved and expanded by our proposal. There are presented nontrivial examples to back up the hypotheses and the value of primary conclusion achieved herein.

Keywords: C.V.M.Space, C.V.N.b-M.Spaces, fixed point, integral equation, fixed point theorems

1. Introduction

F.P.Theory is one of the quite well-accepted fields in mathematics. Theory of Banach [1] fixed point plays the main rule in the execution of solution of nonlinear phenomena. This theory presents the concept of contractive mappings and C.V.M. Space to find out the fixed point of certain function. A different kind of contractive condition that explains the fixed point theorem was presented by Kannan [2] in 1969. The difference between the Banach theorem and the mapping theory of Kannan is that whereas Kannan maps do not always require continuity, Banach mappings do require continuity for contraction. Furthermore, Chaterjea [3] also demonstrated that type of contraction. Numerous academics have improved the Banach [1], F.P.Theorem in a number of different ways (see [4–7]). For additional definitions of contractive sort of mappings, see Rhoades [7]. All of the modifications of the Banach F.P.Theorem are further divided into two categories: those that replace the contractive condition with a more comprehensive one and those that increase or weaken the axioms that define the ground set. In the latter case, a few of these type of M.Spaces have names such as semimetric, pseudometric, quasimetric, K-metric, and bmetric, to name a few. In agreement with all of this, by substituting a typical co-domain for a metric in place \mathbb{R} that is, the set of real numbers, Zhang and Huang [8] originated the cone metric space idea to the literature and enhanced it in different type of metric spaces. By continuing the setting of these ideas, the authors build up a numbers of contractive mappings in F.P. Theorems are proven here in (cone metric spaces). In the setting of cone metric

spaces, other writers have been developed a number of significant fixed point results (see [9, 10]). For the in-depth research study, the interested scholar may go along the study on cone M.Spaces, which was introduced by mathematics scholars named Aleksic et al. [11].

The well-known property in cone M.Spaces, F.P.Theorems using rational con-tractions cannot be extended or rendered irrelevant. Azam et al. [4] developed the concept of C.V.M.Spaces to get around this issue and set up appropriate constraints for a pair of maps to the validity of contractive type inequalities including rational expressions in common fixed points [12, 13]. It is fascinating to note that a specific class of cone metric spaces includes C.V.M.Space. A division ring is not the basis for the definition of a cone metric, which is based on the ground set, that is, Banach space. As a result, many division-related clearly showing that we cannot extrapolated them into a cone metric spaces. The rational inequality was introduced in C.V.M.Spaces and thus new results are proved over there (see, [4, 14–16]). Along these lines, in 1993 Czerwik [17] proposed the notion of b-metric space and Branciari [18], made up a change in triangular inequality to create the idea of rectangle metric space. In addition to complex-valued metric spaces, Rao [19] also proposed the concept of fixed point results on complex-valued b-metric spaces. Every C.V.b-M.Space, however, is a cone b-metric space over Banach algebra \mathbb{C} , where the cone is normal and the normalcy coefficient is $\kappa = 1$. and where the interior of the cone is not empty (i.e., solid cone). With references to C.V.b – M.Spaces, numerous authors have shown fixed point solutions for various maps that satisfy rational inequalities in the paragraphs that follow (see, for illustration, [19–21]).

Inspired by the concepts provided in [4, 17, 18], we define new F.P.Theorems for maps under particular rational constructive inequalities and introduce the notion of C.V.N.b-M.Spaces. The publications cited above, as well as a few others in the related literature, are improved and expanded upon by our idea. Our proposal enhances and expands upon the aforementioned papers as well as a few other sections of related literature.

2. Preliminaries

In this section, we review a few key ideas that are essential to present of our the primary conclusion. [4] Let \mathbb{C} represent the number system of complex numbers and let $p_1, p_2 \in \mathbb{C}$. The definition of the partial order on \mathbb{C} is:

$p_1 \leq p_2, \Leftrightarrow \text{Re}(p_1) \leq \text{Re}(p_2)$ and $\text{Im}(p_1) \leq \text{Im}(p_2)$. This implies that $p_1 \leq p_2$ if one of below follows:

- $[(C_i)] \text{Re}(p_1) = \text{Re}(p_2), \quad \text{Im}(p_1) < \text{Im}(p_2),$
- $[(C_{ii})] \text{Re}(p_1) < \text{Re}(p_2), \quad \text{Im}(p_1) = \text{Im}(p_2),$
- $[(C_{iii})] \text{Re}(p_1) < \text{Re}(p_2), \quad \text{Im}(p_1) < \text{Im}(p_2),$
- $[(C_{iv})] \text{Re}(p_1) = \text{Re}(p_2), \quad \text{Im}(p_1) = \text{Im}(p_2).$

[4] Let \mathbb{M} be a set with function $\psi_g : \mathbb{M} \times \mathbb{M} \rightarrow \mathbb{C}$, which meet the below conditions:

$$\begin{aligned}(C_i) \quad & 0 \leq \psi_g(u, v) \\ (C_{ii}) \quad & \psi_g(u, v) = 0 \Leftrightarrow u = v; \\ (C_{iii}) \quad & \psi_g(u, v) = \psi_g(v, u); \\ (C_{iv}) \quad & \psi_g(u, v) \leq \psi_g(u, w) + \psi_g(w, v), \forall u, w, v \in \mathbb{M},\end{aligned}$$

then ψ_g is referred as a C.V.Metric on \mathbb{M} and (\mathbb{M}, ψ_g) is known as C.V.Mspace. Let $\mathbb{M} = M_1 \cup M_2$, where

$$M_1 = \{h \in \mathbb{C} : \operatorname{Re}(h) \geq 0 \text{ and } \operatorname{Im}(h) = 0\}$$

and

$$M_2 = \{h \in \mathbb{C} : \operatorname{Re}(h) = 0 \text{ and } \operatorname{Im}(h) \geq 0\}.$$

Define $d : \mathbb{M} \times \mathbb{M} \rightarrow \mathbb{C}$ as follows:

$$\psi_g(p_1, p_2) = \begin{cases} \frac{2}{3}|u_1 - u_2| + \frac{i}{2}|u_1 - u_2|, & \text{if } p_1, p_2 \in M_1; \\ \frac{1}{2}|v_1 - v_2| + \frac{i}{2}|v_1 - v_2|, & \text{if } p_1, p_2 \in M_2; \\ \left(\frac{1}{2}v_1 + \frac{2}{3}u_2\right) + i\left(\frac{1}{3}v_1 + \frac{1}{2}u_2\right), & \text{if } p_1 \in M_1, p_2 \in M_2, \end{cases}$$

where $p_1 = u_1 + iv_1$ and $p_2 = u_2 + iv_2$. Then, (\mathbb{M}, ψ_g) is a C.V.M.Space. [4] Let $\mathbb{M} \neq \phi$ be a set with mapping: $\psi_g : \mathbb{M} \times \mathbb{M} \rightarrow \mathbb{C}$ which meet the below requirements:

$$\begin{aligned}(C_i) \quad & 0 \leq \psi_g(u, v) \quad (C_{ii}) \quad \psi_g(u, v) = 0 \Leftrightarrow u = v; \\ (C_{iii}) \quad & \psi_g(u, v) = \psi_g(v, u); \\ (C_{iii}) \quad & d(u, v) \leq \tau[d(u, w) + d(w, v)], \forall \text{ where } \tau \geq 1, u, w, v \in \mathbb{M},\end{aligned}$$

then ψ_g is known as a C.V.b-Metric on \mathbb{M} , and the (\mathbb{M}, ψ_g) is said to be a C.V. b – M.Space [19]. Let $\mathbb{M} = [0, 1]$. Define a function: $\psi_g : \mathbb{M} \times \mathbb{M} \rightarrow \mathbb{C}$ by

$$\psi_g(u, v) = |u - v|^2 + i|u - v|^2$$

for all $u, v \in \mathbb{M}$. Then (\mathbb{M}, ψ_g) is a C.V.b – M.Space with $\mu = 2$. [4] Let $\mathbb{M} \neq \phi$ a set with $\theta : \mathbb{M} \times \mathbb{M} \rightarrow [1, \infty)$ be a mapping and the function: $\psi_g : \mathbb{M} \times \mathbb{M} \rightarrow \mathbb{C}$ obeying the below conditions:

$$\begin{aligned}(C_i) \quad & 0 \leq \psi_g(u, v) \\ (C_{ii}) \quad & \psi_g(u, v) = 0 \Leftrightarrow u = v; \\ (C_{iii}) \quad & \psi_g(u, v) = \psi_g(v, u); \\ (C_{iv}) \quad & \psi_g(u, v) \leq \vartheta(u, v)[\psi_g(u, w) + \psi_g(w, v)], \forall u, w, v \in \mathbb{M},\end{aligned}$$

then ψ_g is called C.V.b-Metric on \mathbb{M} , and the (\mathbb{M}, ψ_g) is known to be a C.V.b – M. Space [4]. Let $\mathbb{M} \neq C([r, s], \mathbb{R})$, where $C([r, s])$ represents a set of continuous mappings which are real-valued, defined on $[r, s]$ and a map $\psi_g : \mathbb{M} \times \mathbb{M} \rightarrow [1, \infty)$ is

$$\vartheta(u, v) = |u(z)| + |v(z)| + 2.$$

Also, define $\psi_{\vartheta} : \mathbb{M} \times \mathbb{M} \rightarrow \mathbb{C}$ by

$$\psi_{\vartheta}(u, v) = \max_{z \in [r, s]} |u(z) - v(z)|^2.$$

Then, $(\mathbb{M}, \psi_{\vartheta})$ is C.V.b – M.Space. Let $\mathbb{M} \neq \phi$ be a set, $\vartheta_0, \vartheta : \mathbb{M} \times \mathbb{M} \rightarrow [0, \infty)$ be defined by

$$\vartheta(u, v) = \vartheta_0(u, v) + \mu,$$

for all $u, v \in \mathbb{M}$, and $\mu \geq 1$. Define a mapping: $\psi_{\vartheta} : \mathbb{M} \times \mathbb{M} \rightarrow \mathbb{C}$, if for all $u, v, w \in \mathbb{M}$, the following assertion are valid.

$$\begin{aligned} (C_i) \quad & 0 \leq \psi_{\vartheta}(u, v) \\ (C_{ii}) \quad & \psi_{\vartheta}(u, v) = 0 \text{ if and only if } u = v; \\ (C_{iii}) \quad & \psi_{\vartheta}(u, v) = \psi_{\vartheta}(v, u); \\ (C_{iii}) \quad & \psi_{\vartheta}(u, v) \leq \vartheta(u, v)[\psi_{\vartheta}(u, w) + d(w, v)]. \end{aligned}$$

Then ψ_{ϑ} is known as C.V.N.E.b-Metric on \mathbb{M} and the pair $(\mathbb{M}, \psi_{\vartheta})$ is known as a C.V.N.Eb-M.Space.

2.1 Ingredient of C.V.E.b – M.Space

If we setting $\vartheta_0(u, v) = u + v$ and $\mu = 1$, then we have $\vartheta(u, v) = 1 + u + v = \vartheta^*(u, v)$.

Property (C_{iii}) of 2 will be updated by

$$\psi_{\vartheta}(u, v) \leq \vartheta^*(u, v)[\psi_{\vartheta}(u, w) + \psi_{\vartheta}(w, v)].$$

Thus, the definition 2 with respect to that condition becomes C.V extended b-M.Space.

2.2 Ingredient of C.V b-M.Spaces

If we setting $\vartheta_0(u, v) = 0$, then $\vartheta(u, v) = 0 + \mu$, where $\mu \geq 1$, then property (C_{iii}) of 2 is

$$\psi_{\vartheta}(u, v) \leq \tau[\psi_{\vartheta}(u, w) + \psi_{\vartheta}(w, v)].$$

The notion of complex-valued b-metric spaces is thus defined by definition 2 with this aspect.

2.3 Ingredient of C.V.M.Spaces

Similar to that when we put $\mu = 1$, in the preceding definition, then its gives us the “C.V.M.Spaces.”

2.4 Development of ordinary M.Space

By replacing the ground set of the previous one's \mathbb{C} with \mathbb{R} . thus ordinary M. Spaces is defined. [19] Consider a set $\mathbb{M} = [0, 1]$ and let a function $\psi_g : \mathbb{M} \times \mathbb{M} \rightarrow \mathbb{C}$ by

$$\psi_g(u, v) = |u - v|^2 + i|u - v|^2$$

for all $u, v \in \mathbb{M}$. Then, (\mathbb{M}, ψ_g) is a C.V.N.E.b-M.Space and letting $\mu = 2$ and $\vartheta_0(u, v) = 0$.

[16] Let $\mathbb{M} \neq \emptyset$ set with mapping $\vartheta_0, \vartheta : \mathbb{M} \times \mathbb{M} \rightarrow [0, \infty)$ be defined as:

$$\vartheta(u, v) = 1 + u + v, \quad \vartheta_0(u, v) = u + v, \mu = 1$$

Furthermore, let

- $[(C_i)] \psi_g(u, v) = \frac{i}{uv}, \forall u, v \in (0, 1];$
- $[(C_{ii})] \psi_g(u, v) = 0 \Leftrightarrow u = v$ for all $u, v \in [0, 1];$
- $[(C_{iii})] \psi_g(u, 0) = \psi_g(0, u) = \frac{i}{u}$ for all $u \in (0, 1].$

Then, the pair (\mathbb{M}, ψ_g) is a C.V.N.E.b-M.Space. [16] Let $\mathbb{M} = [0, \infty)$. $\vartheta : \mathbb{M} \times \mathbb{M} \rightarrow [0, \infty)$ be a mapping defined by $\vartheta(u, v) = 1 + u + v$ and $\psi_g : \mathbb{M} \times \mathbb{M} \rightarrow \mathbb{C}$ to be used as

$$\psi_g(u, v) = \begin{cases} 0, & \text{whenever } u = v \\ i, & \text{whenever } u \neq v. \end{cases}$$

Then, (\mathbb{M}, ψ_g) is a C.V non-negative extended b -M.Space. A non-empty set \mathbb{M} with E and F be a self-maps then:

(C_i) An element $u \in \mathbb{M}$ is known as F.Point of F if $Fu = u$.

(C_{ii}) An element $u \in \mathbb{M}$ is known as coincidence point of F and E if $Fu = Eu$ and we will refer $t = Tu = Su$ to a point of coincidence of F and E.

(C_{iii}) An element $u \in \mathbb{M}$ is known as C.F.Point of F and E if $u = Fu = Eu$. [4] Let (\mathbb{M}, ψ_g) be a C.V.R.E.b – M.Space and the $\{u_k\}$ sequence in \mathbb{M} is a convergent sequence $\Leftrightarrow |\psi_g(u_k, u)| \rightarrow 0$ as $k \rightarrow \infty$.

[4] Let (\mathbb{M}, ψ_g) be a C.V.N.E.b-M.Space and the $\{u_k\}$ a sequence in \mathbb{M} . This implies that $\{u_k\}$ is a Cauchy sequence $\Leftrightarrow |\psi_g(u_k, u_l)| \rightarrow 0$ as $k, l \rightarrow \infty$.

3. Main results

Our primary results are presented below.

Theorem 1.1 A complete C.V non-negative extended b -M.Space (\mathbb{M}, ψ_g) , with a mapping $\vartheta, \vartheta_i : \mathbb{M} \times \mathbb{M} \rightarrow [0, \infty)$ ($i = 0, 1, 2$ and $\vartheta = \vartheta_0 + \mu, \mu \geq 1$) and $F, E : \mathbb{M} \rightarrow \mathbb{M}$ satisfying the assertions below:

- i. $\vartheta_1 < \vartheta_2$;
- ii. $\lim_{k \rightarrow \infty} \vartheta(u_{k+1}, u_l) \vartheta_1(u_{k+1}, u_{k+2}) + \vartheta_2(u_{k+1}, u_{k+2}) < 1$;
- iii. $\psi_\vartheta(Fu, Ev) \leq \vartheta_1(u, v) \psi_\vartheta(u, v) + \vartheta_2(u, v) \frac{\psi_\vartheta(u, Fu) \psi_\vartheta(v, Ev)}{1 + \psi_\vartheta(u, v)}$.

Then F, E have a unique common fixed point in \mathbb{M} .

Proof: Let $u_0 \in \mathbb{M}$ be any element in \mathbb{M} . Build up a sequence $\{u_k\}$ such that

$$u_{2n+1} = Fu_{2n}, \quad u_{2n+2} = Eu_{2n+1} \quad (1)$$

for all $n \geq 0$. From hypothesis and 1, we get.

$$\begin{aligned} \psi_\vartheta(u_{2k+1}, u_{2k+2}) &= \psi_\vartheta(Fu_{2k}, Eu_{2k+1}) \leq \vartheta_1(u_{2k}, u_{2k+1}) \psi_\vartheta(u_{2k}, u_{2k+1}) \\ &\quad + \frac{\vartheta_2(u_{2k}, u_{2k+1}) \psi_\vartheta(u_{2k}, Fu_{2k}) \psi_\vartheta(u_{2k+1}, Eu_{2k+1})}{1 + \psi_\vartheta(u_{2k}, u_{2k+1})}, \\ &\leq \vartheta_1(u_{2k}, u_{2k+1}) \psi_\vartheta(u_{2k}, u_{2k+1}) + \vartheta_2(u_{2k}, u_{2k+1}) \psi_\vartheta(u_{2k+1}, Eu_{2k+1}) \frac{\psi_\vartheta(u_{2k}, u_{2k+1})}{1 + \psi_\vartheta(u_{2k}, u_{2k+1})}, \\ &\leq \vartheta_1(u_{2k}, u_{2k+1}) \psi_\vartheta(u_{2k}, u_{2k+1}) + \vartheta_2(u_{2k}, u_{2k+1}) \psi_\vartheta(u_{2k+1}, u_{2k+2}). \end{aligned}$$

This implies that

$$(1 - \vartheta_2(u_{2k}, u_{2k+1})) \psi_\vartheta(u_{2k+1}, u_{2k+2}) \leq \vartheta_1(u_{2k}, u_{2k+1}) \psi_\vartheta(u_{2k}, u_{2k+1}).$$

That is,

$$\begin{aligned} \psi_\vartheta(u_{2k+1}, u_{2k+2}) &\leq \frac{\vartheta_1(u_{2k}, u_{2k+1})}{1 - \vartheta_2(u_{2k}, u_{2k+1})} \psi_\vartheta(u_{2k}, u_{2k+1}) \\ &\leq \frac{\vartheta_1(u_{2k}, u_{2k+1})}{1 - \vartheta_2(u_{2k}, u_{2k+1})} \frac{\vartheta_1(u_{2k-1}, u_{2k})}{1 - \vartheta_2(u_{2k-1}, u_{2k})} \psi_\vartheta(u_{2k-1}, u_{2k}) \\ &\leq \frac{\vartheta_1(u_{2k}, u_{2k+1})}{1 - \vartheta_2(u_{2k}, u_{2k+1})} \frac{\vartheta_1(u_{2k-1}, u_{2k})}{1 - \vartheta_2(u_{2k-1}, u_{2k})} \frac{\vartheta_1(u_{2k-2}, u_{2k-1})}{1 - \vartheta_2(u_{2k-2}, u_{2k-1})} \psi_\vartheta(u_{2k-2}, u_{2k-1}). \end{aligned}$$

By following the same step, we have

$$\begin{aligned} (2) \quad \psi_\vartheta(u_{2n+1}, u_{2n+2}) &\leq \frac{\vartheta_1(u_{2n}, u_{2n+1})}{1 - \vartheta_2(u_{2n}, u_{2n+1})} \frac{\vartheta_1(u_{2n-1}, u_{2n})}{1 - \vartheta_2(u_{2n-1}, u_{2n})} \frac{\vartheta_1(u_{2n-2}, u_{2n-1})}{1 - \vartheta_2(u_{2n-2}, u_{2n-1})} \\ (3) \quad &\vdots \\ (4) \quad &\times \frac{\vartheta_1(u_0, u_1)}{1 - \vartheta_2(u_0, u_1)} \psi_\vartheta(u_0, u_1). \end{aligned}$$

Taking $l > k$, we have

$$\begin{aligned}
 \psi_{\vartheta}(u_k, u_l) &\leq \vartheta(u_k, u_l) \psi_{\vartheta}(u_k, u_{k+1}) + \vartheta(u_k, u_l) \vartheta(u_{k+1}, u_l) \psi_{\vartheta}(u_{k+1}, u_{k+2}) + \dots \\
 &\quad + \vartheta(u_k, u_l) \vartheta(u_{k+1}, u_l) \vartheta(u_{k+2}, u_l) \dots \vartheta(u_{l-1}, u_l) \psi_{\vartheta}(u_{l-1}, u_l), \\
 &\leq \vartheta(u_0, u_l) \vartheta(u_1, u_l) \vartheta(u_2, u_l) \dots \vartheta(u_{k-1}, u_l) \vartheta(u_k, u_l) \\
 &\quad \frac{\vartheta_1(u_0, u_1)}{1 - \vartheta_2(u_0, u_1)} \frac{\vartheta_1(u_1, u_2)}{1 - \vartheta_2(u_1, u_2)} \dots \frac{\vartheta_1(u_{k-1}, u_k)}{1 - \vartheta_2(u_{k-1}, u_k)} \psi_{\vartheta}(u_0, u_1) \\
 &\quad + \vartheta(u_0, u_l) \vartheta(u_1, u_l) \vartheta(u_2, u_l) \dots \vartheta(u_k, u_l) \vartheta(u_{k+1}, u_l) \\
 &\quad \frac{\vartheta_1(u_0, u_1)}{1 - \vartheta_2(u_0, u_1)} \frac{\vartheta_1(u_1, u_2)}{1 - \vartheta_2(u_1, u_2)} \dots \frac{\vartheta_1(u_{k+1}, u_{k+2})}{1 - \vartheta_2(u_{k+1}, u_{k+2})} \psi_{\vartheta}(u_0, u_1) \\
 &\quad \cdot \\
 &\quad \cdot \\
 &\quad \cdot \\
 &\quad + \vartheta(u_0, u_l) \vartheta(u_1, u_l) \vartheta(u_2, u_l) \dots \vartheta(u_{l-2}, u_l) \vartheta(u_{l-1}, u_l) \\
 &\quad \frac{\vartheta_1(u_0, u_1)}{1 - \vartheta_2(u_0, u_1)} \frac{\vartheta_1(u_1, u_2)}{1 - \vartheta_2(u_1, u_2)} \dots \frac{\vartheta_1(u_{l-2}, u_{l-1})}{1 - \vartheta_2(u_{l-2}, u_{l-1})} \psi_{\vartheta}(u_0, u_1).
 \end{aligned}$$

This implies

$$\begin{aligned}
 \psi_{\vartheta}(u_k, u_l) &\leq \vartheta(u_k, u_l) \psi_{\vartheta}(u_k, u_{k+1}) + \vartheta(u_k, u_l) \vartheta(u_{k+1}, u_l) \psi_{\vartheta}(u_{k+1}, u_{k+2}) + \dots \\
 &\quad + \vartheta(u_k, u_l) \vartheta(u_{k+1}, u_l) \vartheta(u_{k+2}, u_l) \dots \vartheta(u_{l-1}, u_l) \psi_{\vartheta}(u_{l-1}, u_l), \\
 &\leq \psi_{\vartheta}(u_0, u_l) [\vartheta(u_0, u_l) \vartheta(u_1, u_l) \vartheta(u_2, u_l) \dots \vartheta(u_{k-1}, u_l) \vartheta(u_k, u_l) \\
 &\quad \frac{\vartheta_1(u_0, u_1)}{1 - \vartheta_2(u_0, u_1)} \frac{\vartheta_1(u_1, u_2)}{1 - \vartheta_2(u_1, u_2)} \dots \frac{\vartheta_1(u_{k-1}, u_k)}{1 - \vartheta_2(u_{k-1}, u_k)} \\
 &\quad + \vartheta(u_0, u_l) \vartheta(u_1, u_l) \vartheta(u_2, u_l) \dots \vartheta(u_k, u_l) \vartheta(u_{k+1}, u_l) \\
 &\quad \frac{\vartheta_1(u_0, u_1)}{1 - \vartheta_2(u_0, u_1)} \frac{\vartheta_1(u_1, u_2)}{1 - \vartheta_2(u_1, u_2)} \dots \frac{\vartheta_1(u_{k+1}, u_{k+2})}{1 - \vartheta_2(u_{k+1}, u_{k+2})} \\
 &\quad \cdot \\
 &\quad \cdot \\
 &\quad \cdot \\
 &\quad + \vartheta(u_0, u_l) \vartheta(u_1, u_l) \vartheta(u_2, u_l) \dots \vartheta(u_{l-2}, u_l) \vartheta(u_{l-1}, u_l) \\
 &\quad \frac{\vartheta_1(u_0, u_1)}{1 - \vartheta_2(u_0, u_1)} \frac{\vartheta_1(u_1, u_2)}{1 - \vartheta_2(u_1, u_2)} \dots \frac{\vartheta_1(u_{l-2}, u_{l-1})}{1 - \vartheta_2(u_{l-2}, u_{l-1})}].
 \end{aligned}$$

Since, $\lim_{k \rightarrow \infty} \vartheta(u_{k+1}, u_l) \vartheta_1(u_{k+1}, u_{k+2}) + \vartheta_2(u_{k+1}, u_{k+2}) < 1$, thus the series. $\sum_{k=0}^{\infty} \prod_{i=0}^k \frac{\vartheta(u_i, u_l) \vartheta_1(u_i, u_{i+1})}{1 - \vartheta_1(u_i, u_{i+1})}$ converges under the ratio test. For each $l \in \mathbb{N}$, let

$$S_{l-1} = \sum_{j=0}^{l-1} \prod_{i=0}^j \frac{\vartheta(u_i, u_l) \vartheta_1(u_i, u_{i+1})}{1 - \vartheta_1(u_i, u_{i+1})}, S_n = \sum_{j=0}^n \prod_{i=0}^j \frac{\vartheta(u_i, u_l) \vartheta_1(u_i, u_{i+1})}{1 - \vartheta_1(u_i, u_{i+1})}.$$

Conceder $l > k$, for which the above inequality becomes.

$$\psi_{\vartheta}(u_k, u_l) \leq \psi_{\vartheta}(u_0, u_l) [S_{l-1} - S_k].$$

Which means that,

$$|\psi_{\vartheta}(u_k, u_l)| \leq |\psi_{\vartheta}(u_0, u_l)| [S_{l-1} - S_k]. \quad (2)$$

Taking $k \rightarrow \infty$ in (5), we observe that $\{u_k\}$ becomes a Cauchy sequence. Since \mathbb{M} is complete, thus we find out $u \in \mathbb{M}$ such that $u_k \rightarrow u (k \rightarrow \infty)$.

To verify that $Fu = u$, take.

$$\begin{aligned} \psi_\vartheta(u, Fu) &\leq \psi_\vartheta(u, u_{2k+2}) + \psi_\vartheta(u_{2k+2}, Fu) = \psi_\vartheta(u, u_{2k+2}) + \psi_\vartheta(Eu_{2k+1}, Fu) = \\ &\psi_\vartheta(u, u_{2k+2}) + \psi_\vartheta(Fu, Eu_{2k+1}) \leq \psi_\vartheta(u, u_{2k+2}) + \vartheta_1(u, u_{2k+1})\psi_\vartheta(u, u_{2k+1}) + \\ &\vartheta_2(u, u_{2k+1})\psi_\vartheta(u, Fu)\psi_\vartheta(u_{2k+1}, Eu_{2k+1}). \end{aligned}$$

Then, it becomes:

$|\psi_\vartheta(u, Su)| \leq |\psi_\vartheta(u, u_{2k+2})| + |\vartheta_1(u, u_{2k+1})\psi_\vartheta(u, u_{2k+1})| + \left| \frac{\vartheta_2(u, u_{2k+1})\psi_\vartheta(u, Fu)\psi_\vartheta(u_{2k+1}, Eu_{2k+1})}{1 + \psi_\vartheta(u, u_{2k+1})} \right|$. By taking $k \rightarrow \infty$ implies that, $|\psi_\vartheta(u, Fu)| \leq 0$. Then, we conclude that, $u = Fu$. repeating the same way, we can show that $u = Eu$. Hence, F and E have C.F.Point.

3.1 Uniqueness

To verify the uniqueness of u , let us assume that F and E have another F . Point u^* with $u \neq u^*$. Then,

$\psi_\vartheta(u, u^*) = \psi_\vartheta(Fu, Eu^*) \leq \vartheta_1(u, u^*)\psi_\vartheta(u, u^*) + \frac{\vartheta_2(u, u^*)\psi_\vartheta(u, Fu)\psi_\vartheta(u^*, Tu^*)}{1 + \psi_\vartheta(u, u^*)}$, which gives

$$|\psi_\vartheta(u, u^*)| \leq \vartheta_1(u, u^*)|\psi_\vartheta(u, u^*)| + \left| \frac{\vartheta_2(u, u^*)\psi_\vartheta(u, Fu)\psi_\vartheta(u^*, Eu^*)}{1 + \psi_\vartheta(u, u^*)} \right|$$

and

$$(1 - \vartheta_1(u, u^*))\psi_\vartheta(u, u^*) \leq 0,$$

which verified that $u = u^*$. Hence, F and E have a unique F.Point.

A complete C.V.N.E.b-M.Space $(\mathbb{M}, \psi_\vartheta)$, with $\vartheta, \vartheta_i : \mathbb{M} \times \mathbb{M} \rightarrow [0, \infty)$ ($i = 0, 1, 2$ and $\vartheta = \vartheta_0 + \mu, \mu \geq 1$) and $F : \mathbb{M} \rightarrow \mathbb{M}$ be a function with the below conditions:

- i. $\vartheta_1 < \vartheta_2$;
- ii. $\lim_{k \rightarrow \infty} \vartheta(u_{k+1}, u_m)\vartheta_1(u_{k+1}, u_{k+2}) + \vartheta_2(u_{k+1}, u_{k+2}) < 1$;
- iii. $\psi_\vartheta(Fu, Fv) \leq \vartheta_1(u, v)\psi_\vartheta(u, v) + \vartheta_2(u, v)\frac{\psi_\vartheta(u, Fu)\psi_\vartheta(v, Fv)}{1 + \psi_\vartheta(u, v)}$.

Hence, this leads us to the unique F.Point of F .

Proof: Letting $F = E$ in theorem 1.1.

Let us take a result from ([4], Theorem 4) (Author A. Azam et al.) and consider \mathbb{M} as a complete C.V.M.Space and $F, E : \mathbb{M} \rightarrow \mathbb{M}$. If F and E following the below inequality:

$$\psi_\vartheta(Fu, Ev) \leq \alpha\psi_\vartheta(u, v) + \frac{\kappa\psi_\vartheta(u, Fu)\psi_\vartheta(v, Ev)}{1 + \psi_\vartheta(u, v)}$$

for all $u, v \in \mathbb{M}$, where non-negative real numbers α, κ , with $\alpha + \kappa < 1$. Then F and E have a C.F.Point in \mathbb{M} .

Proof: Put $\vartheta_1(u, v) = \alpha$ and $\vartheta_2(u, v) = \kappa$ in 3.

Theorem 1.2 Let \mathbb{M} be a complete C.V.N.E.b-M.Space, $\vartheta, \vartheta_i : \mathbb{M} \times \mathbb{M} \rightarrow [0, \infty)$ ($i = 0, 1, 2$ and $\vartheta = \vartheta_0 + \mu, \mu \geq 1$) and $E : \mathbb{M} \rightarrow \mathbb{M}$ be mappings verifying the following assertions:

- i. $\vartheta_1 < \vartheta_2$;
- ii. $\lim_{k \rightarrow \infty} \vartheta(u_{k+1}, u_l) \vartheta_1(u_{k+1}, u_{k+2}) + \vartheta_2(u_{k+1}, u_{k+2}) < 1$;
- iii. $\psi_\vartheta(E^k u, E^k v) \leq \vartheta_1(u, v) \psi_\vartheta(u, v) + \vartheta_2(u, v) \frac{\psi_\vartheta(u, E^k u) \psi_\vartheta(v, E^k v)}{1 + \psi_\vartheta(u, v)}$.

Then E has a unique F.Point in \mathbb{M} .

Proof: From 3, we observe that E^k has one and only one F.Point u , such that, $E^k u = u$.

Thus, the result is follow below:

$$\begin{aligned} \psi_\vartheta(Eu, u) &= \psi_\vartheta(EE^k u, E^k u) = \psi_\vartheta(E^k Eu, E^k u) \\ &\leq \vartheta_1(Eu, u) \psi_\vartheta(Eu, u) + \frac{\vartheta_2(Eu, u) \psi_\vartheta(Eu, E^n Eu) \psi_\vartheta(u, E^n u)}{1 + \psi_\vartheta(Eu, u)}, \\ &\leq \vartheta_1(Eu, u) \psi_\vartheta(Eu, u) + \frac{\vartheta_2(Eu, u) \psi_\vartheta(Eu, E^n Eu) \psi_\vartheta(u, u)}{1 + \psi_\vartheta(Eu, u)} \\ &= \vartheta_1(Eu, u) \psi_\vartheta(Eu, u), \end{aligned}$$

these give us, $(1 - \vartheta_1(Eu, u)) \psi_\vartheta(Eu, u) \leq 0$. As a result, E has a unique F.Point. Let $\mathbb{M} = C([1, 3], \mathbb{R})$, $d > 0$ and for any $u, v \in \mathbb{M}$, take

$$\psi_\vartheta(u, v) = \max_{z \in [1, 3]} |u(z) - v(z)| \sqrt{1 + d^2} e^{i \tan^{-1} d}.$$

Define $E : \mathbb{M} \rightarrow \mathbb{M}$ by

$$E(r(z)) = 4 + \int_1^z (u(r) + u^2) e^{r-1} dr, z \in [1, 3].$$

Then, for every $u, v \in \mathbb{M}$,

$$\begin{aligned} \psi_\vartheta(Eu, Ev) &= \max_{t \in [1, 3]} |E(u(z)) - E(v(z))| \sqrt{1 + d^2} e^{i \tan^{-1} d}, \\ &\leq \vartheta_1(u, v) \int_1^3 \max_{z \in [1, 3]} |u(r) - v(r)| e^2 \sqrt{1 + d^2} e^{i \tan^{-1} d} dr, \\ &\leq 2 \vartheta_1(u, v) e^2 \psi_\vartheta(u, v). \end{aligned}$$

$$e^{2k} \vartheta_1^k(u, v) \frac{2^k}{k!} = \begin{cases} 109, & \text{if } k = 2 \\ 1987, & \text{if } k = 4 \\ 1.31, & \text{if } k = 37 \\ 0.53, & \text{if } k = 38. \end{cases}$$

By using a typical calculation, we have

$$\psi_g(E^k u, E^k v) \leq e^{2k} \vartheta_1^k(u, v) \frac{2^k}{k!} \psi_g(u, v).$$

Thus, for $\vartheta_1(u, v) = 0.53, \vartheta_2(u, v) = 0, k = 38$, all the assertions of 1.2 are verified and thus, E has a unique F.Point, this implies that the integral equation has unique solution:

$$u(z) = 4 + \int_1^z (u(r) + r^2) e^{r-1} dr, z \in [1, 3],$$

or the differential equation:

$$\frac{du}{dz} = (u + z^2) e^{z-1}, z \in [1, 3], u(1) = 4.$$

4. Conclusions

Inspired by the concepts provided in [4, 17, 18], we define new F.P.Theorems for maps under particular rational constructive inequalities and introduce the notion of C.V.N.b-M.Spaces. The publications cited above, as well as a few others in the related literature, are improved and expanded upon by our idea. Our proposal enhances and expands upon the aforementioned papers as well as a few other sections of related literature.

Conflict of interest

The authors declare that they have no conflict of interests.

Abbreviations

C.C.V.R.b-M.S	Complete Complex Valued Rectangular Extended b-Metric
F.P.Theorem	Fixed Point Theorem
C.C.V.b-M.S	Complete Complex Valued Extended b-Metric Spaces
C.F.Point	Common Fixed Point
M.Spaces	Metric Spaces
F.P	Fixed Point

Author details

Naimat Ullah^{1,2*†}, Mohammed Shehu Shagari^{2†}, Tahir Ahmad Khan^{3†},
Aziz Ullah Khan^{4†} and Muhammad Atta Ullah Khan^{2,5†}

1 Department of Mathematics, University of Mianwali, Mianwali, Pakistan

2 Department of Mathematics, Faculty of Physical Sciences Ahmadu Bello University,
Zaria, Nigeria

3 Department of Mathematics, University of Mianwali, Mianwali, Pakistan


4 Government Graduate College, Mianwali, Pakistan

5 University of the Punjab, Quaid-e-Azam Campus, Lahore, Pakistan

*Address all correspondence to: naimat0347@gmail.com

† These authors contributed equally.

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Banach S. Sur les operation dans les ensembles abstraits et applications aux equations integrales. *Fund. Math.* (French). 1922;**3**:133-181
- [2] Kannan R. Some results on fixed points. *Bulletin of the Calcutta Mathematical Society.* 1968;**60**: 71-76
- [3] Chatterjea SK. Fixed-point theorems. *Dokladi na Bolgarskata Akademiya na Naukite.* 1972;**25**(6):727-735
- [4] Azam A, Fisher B, Khan M. Common fixed point theorems in complex valued metric spaces. *Numerical Functional Analysis and Optimization.* 2011;**32**(3): 243-225
- [5] Azam A, Ahmad J, Kumam P. Common fixed point theorems for multi-valued mappings in complex-valued metric spaces. *Journal of Inequalities and Applications.* 2013; **2013**(1):578
- [6] Edelstein M. On fixed and periodic points under contractive mappings. *Journal of the London Mathematical Society.* 1962;**1**(1):74-79
- [7] Rhoades BE. A comparison of various definitions of contractive mappings. *Transactions of the American Mathematical Society.* 1977;**226**:257-290
- [8] Huang LG, Zhang X. Cone metric spaces and fixed point theorems of contractive mappings. *Journal of Mathematical Analysis and Applications.* 2007;**332**(2):1468-1476
- [9] Jankovic S, Golubovic Z, Radenovic S. Compatible and weakly compatible mappings in cone metric spaces. *Mathematical and Computer Modelling.* 2010;**52**(9–10):1728-1738
- [10] Shatanawi W, Rajic VC, Radenovic S, Al-Rawashdeh A. Mizoguchi-Takahashi-type theorems in tvs-cone metric spaces. *Fixed Point Theory and Applications.* 2012; **2012**(1):106
- [11] Aleksic S, Kadelburg Z, Mitrovic ZD, Radenovic S. A new survey: Cone metric spaces. *Journal of the International Mathematical Virtual Institute.* 2019;**9**: 93-121
- [12] Dubey AK. Complex valued-metric spaces and common fixed point theorems under rational contractions. *Journal of Complex analysis.* 2016;**2016**: 9786063
- [13] Kamran T, Samreen M, Ain QU. A generalization of b-metric space and some fixed point theorems. *Mathematics.* 2017;**5**(2):19
- [14] Dubey AK. Common fixed point results for contractive mappings in complex valued b-metric spaces. *Nonlinear Functional Analysis and Applications.* 2015;**20**: 257-268
- [15] Kumar J. Common fixed point theorem for generalized contractive type maps on complex valued b-metric spaces. *International Journal of Mathematical Analysis.* 2015;**9**: 2327-2334
- [16] Ullah N, Shagari MS, Azam A. Fixed point theorems in complex valued extended b-metric spaces. *Moroccan Journal of Pure and Applied Analysis.* 2019;**5**(2):140-163
- [17] Czerwik S. Contraction mappings in b-metric spaces. *Acta Mathematica Informatica Universitatis Ostraviensis.* 1993;**1**(1):5-11

[18] Mitrovic ZD, Radenovic S. On Meir-Keeler contraction in Branciari b-metric spaces. *Topological Algebra and Its Applications*. 2018;**9**:96-104

[19] Rao P, Swamy R, Prasad JR. A common fixed point theorem in complex valued b-metric spaces. *Bulletin of Mathematics and Statistics Research*. 2013;**1**:1-8

[20] Aiman AM. Some common fixed point theorems in complex valued-metric spaces. *The Scientific World Journal*. 2014;**2014**:1-7

[21] Azam A, Mehmood N. Multivalued fixed point theorems in tvs-cone metric spaces. *Fixed Point Theory and Applications*. 2013;**2013**(1):184

Fixed Point Theorems in Complex-Valued Double Controlled Metric-Like Spaces

*Mohammad S.R. Chowdhury, Muhammad Suhail Aslam,
Musawa Yahya Almusawa and Muhammad Imran Asjad*

Abstract

In this chapter, we have discussed the concept of complex-valued double controlled metric-like spaces. These results extend the corresponding results about complex-valued double controlled metric-type spaces. As applications, we have proved some complex-valued fixed point theorems in this space and have proved the existence and uniqueness of the solution of a Fredholm type integral equation. Finally, some examples are also given in favor of our given results.

Keywords: fixed point, complex-valued, double controlled, metric-like spaces, Fredholm type integral equations

1. Introduction

Fixed point theory concept is a widely recognized and big subject of studies in engineering and mathematical sciences. This area is called the aggregate of evaluation including geometry, algebra, and topology. In this field, a big contribution has been done by Banach [1] by introducing notion of contraction mapping due to a complete metric space to find fixed point of the specified function. This classical theorem of Banach [1] has been studied and generalized by means of many researchers in diverse methods (see, for instance, [2–20]), as this principle is also considered as heart of fixed point theory. Classical definition of metric space was generalized by Harandi [21] by introducing the notion of metric-like space. Azam [22] was the first one to introduce the complex-valued metric space, which was generalized by Hosseini and Karizaki [23], who introduced complex-valued metric-like space. Bakhtin [2] and Czerwik [3] generalized metric space by giving the idea of b-metric spaces. Aslam et al. [4] introduced the notion of complex-valued controlled metric-type space. Abdeljawad [7] presented the idea of double controlled metric-type space (DCMLS) that was generalization of Kamran et al. [5] and Mlaiki et al. [6].

Further, let us recall some interesting definitions and results, useful in the introduction of our new concept.

Let \mathbb{C} be the set of complex numbers and $w_1, w_2 \in \mathbb{C}$. Since we cannot compare in usual way two complex numbers let us add to the complex set \mathbb{C} the following partial order \lesssim , known in related literature as lexicographic order.

$w_1 \lesssim w_2$ if and only if $Re(w_1) \leq Re(w_2)$ or $(Re(w_1) = Re(w_2) \text{ and } Im(w_1) \leq Im(w_2))$. Taking into account the previous definition, we have that $w_1 \lesssim w_2$ if one of the next conditions is satisfied:

(P1) $Re(w_1) < Re(w_2)$ and $Im(w_1) < Im(w_2)$;

(P2) $Re(w_1) < Re(w_2)$ and $Im(w_1) = Im(w_2)$;

(P3) $Re(w_1) < Re(w_2)$ and $Im(w_1) > Im(w_2)$;

(P4) $Re(w_1) = Re(w_2)$ and $Im(w_1) < Im(w_2)$.

Let us recall the definition of complex-valued extended b -metric given by N. Ullah et al. in Ref. [9].

Definition [9]

Let \mathbb{X} be a non-empty set and $\vartheta : \mathbb{X} \times \mathbb{X} \rightarrow [1, \infty)$. The function $h_{eb} : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{C}$ is said to be complex-valued extended b -metric if the following conditions are satisfied:

$$(CEB_1) \ 0 \lesssim h_{eb}(p, q) \text{ and } h_{eb}(p, q) = 0 \text{ if and only if } p = q,$$

$$(CEB_2) h_{eb}(p, q) = h_{eb}(q, p),$$

$$(CEB_3) h_{eb}(p, r) \lesssim \vartheta(p, r) [h_{eb}(p, q) + h_{eb}(q, r)],$$

for all $p, q, r \in \mathbb{X}$. A pair (\mathbb{X}, h_{eb}) is called a complex-valued extended b -metric space.

Mlaiki et al. [6] generalized the notion of b -metric spaces as follows.

Definition [6]

Given $\varrho : \mathbb{X} \times \mathbb{X} \rightarrow [1, \infty)$, where \mathbb{X} is nonempty. Let $h_c : \mathbb{X} \times \mathbb{X} \rightarrow [0, \infty)$. Suppose that

$$(CMT_1) h_c(p, q) = 0 \text{ if and only if } p = q,$$

$$(CMT_2) h_c(p, q) = h_c(q, p),$$

$$(CMT_3) h_c(p, q) \leq \varrho(p, r) h_c(p, r) + \varrho(r, q) h_c(r, q),$$

for all $p, q, r \in \mathbb{X}$. Then, h_c is called a controlled metric type and (\mathbb{X}, h_c) is called a controlled metric-type space.

Definition [13]

Given non-comparable functions $\varrho, \varsigma : \mathbb{X} \times \mathbb{X} \rightarrow [1, \infty)$, where \mathbb{X} is nonempty.

If $h_{dl} : \mathbb{X}^2 \rightarrow [0, \infty)$ satisfies

$$(DCML_1) h_{dl}(p, q) = 0 \Rightarrow p = q,$$

$$(DCML_2) h_{dl}(p, q) = h_{dl}(q, p),$$

$$(DCML_3) h_{dl}(p, q) \leq \varrho(p, r) h_{dl}(p, r) + \varsigma(r, q) h_{dl}(r, q),$$

for all $p, q, r \in \mathbb{X}$. Then, h_{dl} is called a double controlled metric like by ϱ and ς and (\mathbb{X}, h_{dl}) is called a double controlled metric-like space (DCMLS).

Definition [8]

Given non-comparable functions $\varrho, \varsigma : \mathbb{X} \times \mathbb{X} \rightarrow [1, \infty)$, where \mathbb{X} is nonempty.

If $h_{cdt} : \mathbb{X}^2 \rightarrow [0, \infty)$ satisfies

$$\begin{aligned}(CDCMT_1)h_{cdt}(p, q) &= 0 \Leftrightarrow p = q, \\(CDCMT_2)h_{cdt}(p, q) &= h_{cdt}(q, p), \\(CDCMT_3)h_{cdt}(p, q) &\lesssim \varrho(p, r)h_{cdt}(p, r) + \varsigma(r, q)h_{cdt}(r, q),\end{aligned}$$

for all $p, q, r \in \mathbb{X}$. Then, h_{cdt} is called a complex-valued double controlled metric type by ϱ and ς and (\mathbb{X}, h_{cdt}) is called a complex-valued double controlled metric-type space (CDCMTS).

Recently, Panda et al. [8] presented idea of complex-valued double controlled metric-type space (CDCMTS). Inspired by Panda et al. [8], in this chapter we will present the concept of complex-valued double controlled metric-like space (CDCMLS). Then, two fixed point theorems in CDCMLS are presented. One of them is the Banach contraction principle and the second one is the related Reich type result. The theorems are illustrated with the help of examples. Moreover, an application to prove the existence and the uniqueness of a Fredholm type integral equation is given.

2. Complex-valued double controlled metric-like spaces

In the section, we will introduce our generalization of complex-valued double controlled metric-like spaces (CDCMLS).

Definition

Given non-comparable functions $\varrho, \varsigma : \mathbb{X} \times \mathbb{X} \rightarrow [1, \infty)$, where \mathbb{X} is nonempty. If $h_{cdl} : \mathbb{X}^2 \rightarrow [0, \infty)$ satisfies

$$\begin{aligned}(CDCML_1)h_{cdl}(p, q) &= 0 \Rightarrow p = q, \\(CDCML_2)h_{cdl}(p, q) &= h_{cdl}(q, p), \\(CDCML_3)h_{cdl}(p, r) &\lesssim \varrho(p, q)h_{cdl}(p, q) + \varsigma(q, r)h_{cdl}(q, r),\end{aligned}$$

for all $p, q, r \in \mathbb{X}$. Then, h_{cdl} is called a complex-valued double controlled metric like by ϱ and ς and (\mathbb{X}, h_{cdl}) is called a complex-valued double controlled metric-like space (CDCMLS).

Remark

A complex-valued double controlled metric-type space is also complex-valued double controlled metric-like space in general. The converse is not true in general. Further this is also more generalized form than complex-valued extended b -metric-type space (see Example 2).

Let $\mathbb{X} = \{1, 2, 3\}$. Consider the complex-valued double controlled metric like $h = h_{cdl}$ defined by

$$\begin{aligned}h(1, 1) &= h(2, 2) = 0 \text{ and } h(3, 3) = \frac{i}{2}, \\h(1, 2) &= h(2, 1) = 2 + 4i, h(2, 3) = h(3, 2) = i, h(1, 3) = h(3, 1) = 1 - i.\end{aligned}$$

Take $\varrho, \varsigma : \mathbb{X} \times \mathbb{X} \rightarrow [1, \infty)$ to be symmetric and defined by

$$\begin{aligned}\varrho(1, 1) &= \varrho(2, 2) = \varrho(3, 3) = 1, \quad \varrho(1, 2) = \varrho(2, 1) = \frac{6}{5}, \quad \varrho(2, 3) = \varrho(3, 2) = \frac{8}{5}, \\ \varrho(3, 1) &= \varrho(1, 3) = \frac{151}{100}.\end{aligned}$$

$$\begin{aligned} \varsigma(1, 1) = \varsigma(2, 2) = \varsigma(3, 3) = 1, \quad \varsigma(1, 2) = \varsigma(2, 1) = \frac{6}{5}, \quad \varsigma(2, 3) = \varsigma(3, 2) = \frac{33}{20}, \\ \text{and} \\ \varsigma(3, 1) = \varsigma(1, 3) = \frac{8}{3}. \end{aligned}$$

The conditions $(CDCML_1)$ and $(CDCML_2)$ hold. We will verify $(CDCML_3)$.

Case 1: When $p = q = r = 1$,

$$\begin{aligned} |h(p, r)| &= |h(1, 1)| = 0 \leq 0 = 0 + 0 = \mathfrak{Q}(1, 1) |h(1, 1)| + \varsigma(1, 1) |h(1, 1)| \\ &= \mathfrak{Q}(p, q) |h(p, q)| + \varsigma(q, r) |h(q, r)|. \end{aligned}$$

Case 2: When $p = 2, q = r = 1$,

$$\begin{aligned} |h(p, r)| &= |h(2, 1)| = \sqrt{20} \leq \frac{6}{5} \sqrt{20} = \frac{6}{5} \sqrt{20} + 0 = \mathfrak{Q}(2, 1) |h(2, 1)| + \varsigma(1, 1) |h(1, 1)| \\ &= \mathfrak{Q}(p, q) |h(p, q)| + \varsigma(q, r) |h(q, r)|. \end{aligned}$$

Case 3: When $p = 3, q = r = 1$,

$$\begin{aligned} |h(p, r)| &= |h(3, 1)| = \sqrt{2} \leq \frac{151}{100} \sqrt{2} = \frac{151}{100} \sqrt{20} + 0 = \mathfrak{Q}(3, 1) |h(3, 1)| + \varsigma(1, 1) |h(1, 1)| \\ &= \mathfrak{Q}(p, q) |h(p, q)| + \varsigma(q, r) |h(q, r)|. \end{aligned}$$

Case 4: When $p = r = 1, q = 2$,

$$\begin{aligned} |h(p, r)| &= |h(1, 1)| = 0 \leq \frac{12}{5} \sqrt{20} = \frac{6}{5} \sqrt{20} + \frac{6}{5} \sqrt{20} = \mathfrak{Q}(1, 2) |h(1, 2)| + \varsigma(2, 1) |h(2, 1)| \\ &= \mathfrak{Q}(p, q) |h(p, q)| + \varsigma(q, r) |h(q, r)|. \end{aligned}$$

Case 5: When $p = r = 1, q = 3$,

$$\begin{aligned} |h(p, r)| &= |h(1, 1)| = 0 \leq \frac{1253}{300} \sqrt{2} = \frac{151}{100} \sqrt{2} + \frac{8}{3} \sqrt{2} = \mathfrak{Q}(1, 3) |h(1, 3)| + \varsigma(3, 1) |h(3, 1)| \\ &= \mathfrak{Q}(p, q) |h(p, q)| + \varsigma(q, r) |h(q, r)|. \end{aligned}$$

Case 6: When $p = q = 1, r = 2$,

$$\begin{aligned} |h(p, r)| &= |h(1, 2)| = \sqrt{20} \leq \frac{6}{5} \sqrt{20} = 0 + \frac{6}{5} \sqrt{20} = \mathfrak{Q}(1, 1) |h(1, 1)| + \varsigma(1, 2) |h(1, 2)| \\ &= \mathfrak{Q}(p, q) |h(p, q)| + \varsigma(q, r) |h(q, r)|. \end{aligned}$$

Case 7: When $p = q = 1, r = 3$,

$$\begin{aligned} |h(p, r)| &= |h(1, 3)| = \sqrt{2} \leq \frac{8}{3} \sqrt{2} = 0 + \frac{8}{3} \sqrt{2} = \mathfrak{Q}(1, 1) |h(1, 1)| + \varsigma(1, 3) |h(1, 3)| \\ &= \mathfrak{Q}(p, q) |h(p, q)| + \varsigma(q, r) |h(q, r)|. \end{aligned}$$

Case 8: When $p = q = 2, r = 1$,

$$\begin{aligned} |h(p, r)| &= |h(2, 1)| = \sqrt{20} \leq \frac{6}{5} \sqrt{20} = 0 + \frac{6}{5} \sqrt{20} = \mathfrak{Q}(2, 2) |h(2, 2)| + \varsigma(2, 1) |h(2, 1)| \\ &= \mathfrak{Q}(p, q) |h(p, q)| + \varsigma(q, r) |h(q, r)|. \end{aligned}$$

Case 9: When $p = 2, q = 3, r = 1$,

$$\begin{aligned} |h(p, r)| &= |h(2, 1)| = \sqrt{20} \leq \frac{24 + 40\sqrt{2}}{15} = \frac{8}{5} + \frac{8}{3}\sqrt{2} = \varrho(2, 3)|h(2, 3)| + \varsigma(3, 1)|h(3, 1)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 10: When $p = 3, q = 2, r = 1$,

$$\begin{aligned} |h(p, r)| &= |h(3, 1)| = \sqrt{2} \leq \frac{8 + 6\sqrt{20}}{5} = \frac{8}{5} + \frac{6}{5}\sqrt{20} = \varrho(3, 2)|h(3, 2)| + \varsigma(2, 1)|h(2, 1)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 11: When $p = q = 3, r = 1$,

$$\begin{aligned} |h(p, r)| &= |h(3, 1)| = \sqrt{2} \leq \frac{3 + 16\sqrt{2}}{6} = \frac{1}{2} + \frac{8}{3}\sqrt{2} = \varrho(3, 3)|h(3, 3)| + \varsigma(3, 1)|h(3, 1)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 12: When $p = 1, q = r = 2$,

$$\begin{aligned} |h(p, r)| &= |h(1, 2)| = \sqrt{20} \leq \frac{6}{5}\sqrt{20} = \frac{6}{5}\sqrt{2} + 0 = \varrho(1, 2)|h(1, 2)| + \varsigma(2, 2)|h(2, 2)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 13: When $p = 1, q = 3, r = 2$,

$$\begin{aligned} |h(p, r)| &= |h(1, 2)| = \sqrt{20} \leq \frac{453\sqrt{2} + 800}{300} = \frac{151}{100}\sqrt{2} + \frac{8}{3} = \varrho(1, 3)|h(1, 3)| + \varsigma(3, 2)|h(3, 2)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 14: When $p = r = 2, q = 1$,

$$\begin{aligned} |h(p, r)| &= |h(2, 2)| = 0 \leq \frac{12}{5}\sqrt{20} = \frac{6}{5}\sqrt{20} + \frac{6}{5}\sqrt{20} = \varrho(2, 1)|h(2, 1)| + \varsigma(1, 2)|h(1, 2)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 15: When $p = q = r = 2$,

$$\begin{aligned} |h(p, r)| &= |h(2, 2)| = 0 \leq 0 = 0 + 0 = \varrho(2, 2)|h(2, 2)| + \varsigma(2, 2)|h(2, 2)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 16: When $p = r = 2, q = 3$,

$$\begin{aligned} |h(p, r)| &= |h(2, 2)| = 0 \leq \frac{13}{4} = \frac{8}{5} + \frac{33}{20} = \varrho(2, 3)|h(2, 3)| + \varsigma(3, 2)|h(3, 2)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 17: When $p = 3, q = 1, r = 2$,

$$\begin{aligned} |h(p, r)| &= |h(3, 2)| = 1 \leq \frac{151\sqrt{2} + 120\sqrt{20}}{100} = \frac{151}{100}\sqrt{2} + \frac{6}{5}\sqrt{20} \\ &= \varrho(3, 1)|h(3, 1)| + \varsigma(1, 2)|h(1, 2)| = \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 18: When $p = 3, q = r = 2$,

$$\begin{aligned} |h(p, r)| &= |h(3, 2)| = 1 \leq \frac{8}{5} = \frac{8}{5} + 0 = \varrho(3, 2)|h(3, 2)| + \varsigma(2, 2)|h(2, 2)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 19: When $p = q = 3, r = 2$,

$$\begin{aligned} |h(p, r)| &= |h(3, 2)| = 1 \leq \frac{43}{20} = \frac{1}{2} + \frac{33}{20} = \varrho(3, 3)|h(3, 3)| + \varsigma(3, 2)|h(3, 2)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 20: When $p = 1, q = 2, r = 3$,

$$\begin{aligned} |h(p, r)| &= |h(1, 3)| = \sqrt{2} \leq \frac{24\sqrt{20} + 33}{20} = \frac{6}{5}\sqrt{20} + \frac{33}{20} = \varrho(1, 2)|h(1, 2)| + \varsigma(2, 3)|h(2, 3)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 21: When $p = 1, q = r = 3$,

$$\begin{aligned} |h(p, r)| &= |h(1, 3)| = \sqrt{2} \leq \frac{151\sqrt{2} + 50}{100} = \frac{151}{100}\sqrt{2} + \frac{1}{2} = \varrho(1, 3)|h(1, 3)| + \varsigma(3, 3)|h(3, 3)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 22: When $p = 2, q = 1, r = 3$,

$$\begin{aligned} |h(p, r)| &= |h(2, 3)| = 1 \leq \frac{18\sqrt{20} + 40\sqrt{2}}{15} = \frac{6}{5}\sqrt{20} + \frac{8}{3}\sqrt{2} = \varrho(2, 1)|h(2, 1)| + \varsigma(1, 3)|h(1, 3)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 23: When $p = q = 2, r = 3$,

$$\begin{aligned} |h(p, r)| &= |h(2, 3)| = 1 \leq \frac{33}{20} = 0 + \frac{33}{20} = \varrho(2, 2)|h(2, 2)| + \varsigma(2, 3)|h(2, 3)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 24: When $p = 2, q = r = 3$,

$$\begin{aligned} |h(p, r)| &= |h(2, 3)| = 1 \leq \frac{21}{10} = \frac{8}{5} + \frac{1}{2} = \varrho(2, 3)|h(2, 3)| + \varsigma(3, 3)|h(3, 3)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 25: When $p = r = 3, q = 1$,

$$\begin{aligned} |h(p, r)| &= |h(3, 3)| = \frac{1}{2} \leq \frac{1253}{300}\sqrt{2} = \frac{151}{100}\sqrt{2} + \frac{8}{3}\sqrt{2} = \varrho(3, 1)|h(3, 1)| + \varsigma(1, 3)|h(1, 3)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 26: When $p = r = 3, q = 2$,

$$\begin{aligned} |h(p, r)| &= |h(3, 3)| = \frac{1}{2} \leq \frac{13}{4} = \frac{8}{5} + \frac{33}{20} = \varrho(3, 2)|h(3, 2)| + \varsigma(2, 3)|h(2, 3)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Case 27: When $p = q = r = 3$,

$$\begin{aligned} |h(p, r)| &= |h(3, 3)| = \frac{1}{2} \leq 1 = \frac{1}{2} + \frac{1}{2} = \varrho(3, 3)|h(3, 3)| + \varsigma(3, 3)|h(3, 3)| \\ &= \varrho(p, q)|h(p, q)| + \varsigma(q, r)|h(q, r)|. \end{aligned}$$

Thus, $h = h_{cdl}$ is complex-valued double controlled metric-like space (CDCMLS).
 But when $p = 2, q = 3, r = 1$,

$$\begin{aligned} |h(p, r)| &= |h(2, 1)| = \sqrt{20} \leq \frac{6}{5} (1 + \sqrt{2}) = \frac{6}{5} [1 + \sqrt{2}] = \vartheta(2, 1)[|h(2, 3)| + |h(3, 1)|] \\ &= \vartheta(p, q)[|h(p, r)| + |h(q, r)|]. \end{aligned}$$

Thus, $h = h_{cdl}$ is not a complex-valued extended b -metric type for the function ϑ .
 Let (\mathbb{X}, h_{cdl}) be a complex-valued double controlled metric-like space (CDCMLS) by one or two functions.

1. The sequence $\{p_n\}$ is convergent to some p in \mathbb{X} , if for each positive ε , there is some integer Z_ε such that $h_{cdl}(p_n, p) < \varepsilon$ for each $n \geq Z_\varepsilon$. It is written as $\lim_{n \rightarrow \infty} p_n = p$.
2. The sequence $\{p_n\}$ is said Cauchy, if for every $\varepsilon > 0$, $h_{cdl}(p_n, p_m) < \varepsilon$ for all $m, n \geq Z_\varepsilon$, where Z_ε is some integer.
3. (\mathbb{X}, h_{cdl}) is said complete if every Cauchy sequence is convergent.

Let (\mathbb{X}, h_{cdl}) be a complex-valued double controlled metric-like space (CDCMLS) by either one function or two functions—for $p \in \mathbb{X}$ and $l > 0$.

- i. We define $\mathcal{B}(p, l)$ as

$$\mathcal{B}(p, l) = \{y \in \mathbb{X}, h_{cdl}(p, y) < l\}.$$

- ii. The self-map Υ on \mathbb{X} is said to be continuous at p in \mathbb{X} if for all $\delta > 0$, there exists $l > 0$ such that

$$\Upsilon(\mathcal{B}(p, l)) \subseteq \mathcal{B}(\Upsilon p, \delta).$$

Note that if Υ is continuous at p in (\mathbb{X}, h_{cdl}) , then $p_n \rightarrow p$ implies that $\Upsilon p_n \rightarrow \Upsilon p$ when n tends to ∞ .

One can prove the following lemmas for the specific case of CDCMLS, in a similar way as in [11]. Let (\mathbb{X}, h_{cdl}) be a CDCMLS and assume a sequence $\{d_n\}$ in \mathbb{X} . Then $\{d_n\}$ is Cauchy sequence $\Leftrightarrow |h_{cdl}(d_m, d_n)| \rightarrow 0$ as $m, n \rightarrow \infty$, where $m, n \in \mathbb{N}$.

Suppose (\mathbb{X}, h_{cdl}) be a CDCMLS and $\{d_n\}$ be sequence in \mathbb{X} . Then $\{d_n\}$ converges to $d \Leftrightarrow |h_{cdl}(d_n, d)| \rightarrow 0$ as $n \rightarrow \infty$.

Let (\mathbb{X}, h_{cdl}) be a CDCMLS. Then a sequence $\{d_n\}$ in \mathbb{X} is Cauchy sequence, such that $d_m \neq d_n$, whenever $m \neq n$. Then $\{d_n\}$ converges to at most one point.

For a given complex-valued controlled space (\mathbb{X}, h_{cdl}) , the complex-valued double controlled (c.v.dc) metric-like function $h_{cdl} : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{C}$ is continuous, with respect to the partial order “ \lesssim ”.

Consider (\mathbb{X}, h_{cdl}) be a CDCMLS. Limit of every convergent sequence in \mathbb{X} is unique, if the functional $h_{cdl} : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{X}$ is continuous.

3. Fixed point theorems in CDCMLS

In our first theorem of this section, we prove the Banach contraction type theorem in CDCMLS.

Theorem 1.1 Let (\mathbb{X}, h_{cdl}) be a CDCMLS by the functions $\varrho, \varsigma : \mathbb{X} \times \mathbb{X} \rightarrow [1, \infty)$. Suppose that $\Upsilon : \mathbb{X} \times \mathbb{X}$ satisfies

$$h(\Upsilon p, \Upsilon q) \lesssim l h_{cdl}(p, q) \quad (1)$$

for all $p, q \in \mathbb{X}$, where $l \in (0, 1)$. For $p_0 \in \mathbb{X}$, choose $p_n = \Upsilon^n p_0$. Assume that

$$\sup_m \lim_{i \rightarrow \infty} \frac{\varrho(p_{i+1}, p_{i+1})}{\varrho(p_i, p_{i+1})} \varsigma(p_{i+1}, p_m) < \frac{1}{l}. \quad (2)$$

In addition, for each $p \in \mathbb{X}$, suppose that

$$\lim_{n \rightarrow \infty} \varrho(p, p_n) \text{ and } \lim_{n \rightarrow \infty} \varsigma(p_n, p) \text{ exist and are finite.} \quad (3)$$

Then, Υ has a unique fixed point.

Proof: Consider the sequence $\{p_n = \Upsilon^n p_0\}$ in \mathbb{X} that satisfies the hypothesis of the theorem. By using (1), we get

$$h_{cdl}(p_n, p_{n+1}) \lesssim l^n h_{cdl}(p_0, p_1) \text{ for all } n \geq 0. \quad (4)$$

Let n, m be integers such that $n < m$. We have

$$\begin{aligned} h_{cdl}(p_n, p_m) &\lesssim \varrho(p_n, p_{n+1}) h_{cdl}(p_n, p_{n+1}) + \varsigma(p_{n+1}, p_m) h_{cdl}(p_{n+1}, p_m) \\ &\lesssim \varrho(p_n, p_{n+1}) h_{cdl}(p_n, p_{n+1}) + \varsigma(p_{n+1}, p_m) \varrho(p_{n+1}, p_{n+2}) h_{cdl}(p_{n+1}, p_{n+2}) \\ &\quad + \varsigma(p_{n+1}, p_m) \varsigma(p_{n+2}, p_m) h_{cdl}(p_{n+2}, p_m) \\ &\lesssim \varrho(p_n, p_{n+1}) h_{cdl}(p_n, p_{n+1}) + \varsigma(p_{n+1}, p_m) \varrho(p_{n+1}, p_{n+2}) h_{cdl}(p_{n+1}, p_{n+2}) \\ &\quad + \varsigma(p_{n+1}, p_m) \varsigma(p_{n+2}, p_m) \varrho(p_{n+2}, p_{n+3}) h_{cdl}(p_{n+2}, p_{n+3}) \\ &\quad + \varsigma(p_{n+1}, p_m) \varsigma(p_{n+2}, p_m) \varsigma(p_{n+3}, p_m) h_{cdl}(p_{n+3}, p_m) \\ &\lesssim \dots \\ &\lesssim \varrho(p_n, p_{n+1}) h_{cdl}(p_n, p_{n+1}) + \sum_{i=n+1}^{m-2} \left(\prod_{j=n+1}^i \varsigma(p_j, p_m) \right) \varrho(p_i, p_{i+1}) h_{cdl}(p_i, p_{i+1}) \\ &\quad + \prod_{k=n+1}^{m-1} \varsigma(p_k, p_m) h_{cdl}(p_{m-1}, p_m) \\ &\lesssim \varrho(p_n, p_{n+1}) l^n h_{cdl}(p_0, p_1) + \sum_{i=n+1}^{m-2} \left(\prod_{j=n+1}^i \varsigma(p_j, p_m) \right) \varrho(p_i, p_{i+1}) l^i h_{cdl}(p_0, p_1) \\ &\quad + \prod_{i=n+1}^{m-1} \varsigma(p_i, p_m) l^{m-1} h_{cdl}(p_0, p_1) \end{aligned}$$

$$\begin{aligned}
 & \lesssim \varrho(p_n, p_{n+1}) l^m h_{cdl}(p_0, p_1) + \sum_{i=n+1}^{m-2} \left(\prod_{j=n+1}^i \varsigma(p_j, p_m) \right) \varrho(p_i, p_{i+1}) l^i h_{cdl}(p_0, p_1) \\
 & + \left(\prod_{i=n+1}^{m-1} \varsigma(p_i, p_m) \right) l^{m-1} \varrho(p_{m-1}, p_m) h_{cdl}(p_0, p_1) \\
 & = \varrho(p_n, p_{n+1}) l^m h_{cdl}(p_0, p_1) + \sum_{i=n+1}^{m-1} \left(\prod_{j=n+1}^i \varsigma(p_j, p_m) \right) \varrho(p_i, p_{i+1}) l^i h_{cdl}(p_0, p_1) \\
 & \lesssim \varrho(p_n, p_{n+1}) l^m h_{cdl}(p_0, p_1) + \sum_{i=n+1}^{m-1} \left(\prod_{j=0}^i \varsigma(p_j, p_m) \right) \varrho(p_i, p_{i+1}) l^i h_{cdl}(p_0, p_1).
 \end{aligned}$$

We used $(p, q) \geq 1$. Let

$$R_g + \sum_{i=0}^g \left(\prod_{j=0}^i \varsigma(p_j, p_m) \right) \varrho(p_i, p_{i+1}) l^i.$$

Hence, we have

$$h_{cdl}(p_n, p_m) \lesssim h_{cdl}(p_0, p_1) [l^n \varrho(p_n, p_{n+1}) + (R_{m-1}, R_n)]. \quad (5)$$

The ratio test together with (2) implies that the limit of the real number sequence $\{R_n\}$ exists, and so $\{R_n\}$ is Cauchy.

Indeed, the ration test is applied to the term $v_i = \left(\prod_{j=0}^i \varsigma(p_j, p_m) \right) \varrho(p_i, p_{i+1}) \dots$

Letting n, m tend to ∞ in (5) yields

$$\lim_{n, m \rightarrow \infty} h_{cdl}(p_n, p_m) = 0,$$

so the sequence $\{p_n\}$ is Cauchy. Since (\mathbb{X}, h_{cdl}) is a complete double controlled metric-type space, there exists some $\kappa \in \mathbb{X}$ such that

$$\lim_{n \rightarrow \infty} h_{cdl}(p_n, \kappa) = 0.$$

We claim that $\Upsilon\kappa = \kappa$. By (DCML3), we have

$$h(\kappa, p_{n+1}) \lesssim \varrho(\kappa, p_n) h_{cdl}(\kappa, p_n) + \varsigma(p_n, p_{n+1}) h_{cdl}(p_n, p_{n+1}). \quad (6)$$

Using (3) and (6), we get that

$$\lim_{n \rightarrow \infty} h(\kappa, p_{n+1}) = 0. \quad (7)$$

By (1), we have

$$\begin{aligned}
 h(\kappa, \Upsilon\kappa) & \lesssim \varrho(\kappa, p_{n+1}) h_{cdl}(\kappa, p_{n+1}) + \varsigma(p_{n+1}, \Upsilon\kappa) h_{cdl}(p_{n+1}, \Upsilon\kappa) \\
 & \lesssim \varrho(p, p_{n+1}) h_{cdl}(\kappa, p_{n+1}) + l \varsigma(p_{n+1}, \Upsilon\kappa) h_{cdl}(p_n, \kappa).
 \end{aligned}$$

Using (3) and (7), we get at the limit $h_{cdl}(\kappa, \Upsilon\kappa) = 0$, that is, $\Upsilon\kappa = \kappa$. Let ϖ in \mathbb{X} be such that $\Upsilon\eta = \varpi$ and $\kappa \neq \varpi$. We have

$$0 < h_{cdl}(\kappa, \varpi) = h_{cdl}(\Upsilon\kappa, \Upsilon\kappa) \leq lh_{cdl}(\kappa, \varpi).$$

It is a contradiction, so $\kappa = \varpi$. Hence, κ is the unique fixed point of Υ .

The assumption (3) in Theorem 1.1 above can be replaced by the assumptions that the mapping Υ and the complex-valued double controlled metric h are continuous. Indeed, when $p_n \rightarrow \kappa$, then $\Upsilon p_n \rightarrow \Upsilon\kappa$ and hence we have

$$\lim_{n \rightarrow \infty} h_{cdl}(\Upsilon p_n, \Upsilon\kappa) = 0 = \lim_{n \rightarrow \infty} h_{cdl}(\Upsilon p_{n+1}, \Upsilon\kappa) = p(\kappa, \Upsilon\kappa),$$

and hence $\Upsilon\kappa = \kappa$.

Theorem 1.1 is illustrated by the following examples.

We endow $\mathbb{X} = \{1, 2, 3\}$ by the following CDCMLS $h = h_{cdl}$

$$h(1, 1) = h(2, 2) = 0 \text{ and } h(3, 3) = \frac{i}{2},$$

$$h(1, 2) = h(2, 1) = 2 + 4i, \quad h(2, 3) = h(3, 2) = i, \quad h(3, 1) = h(3, 1) = 1 - i.$$

Take $\varrho, \varsigma : \mathbb{X} \times \mathbb{X} \rightarrow [1, \infty)$ to be symmetric and defined by

$$\varrho(1, 1) = \varrho(2, 2) = \varrho(3, 3) = 1, \quad \varrho(1, 2) = \varrho(2, 1) = \frac{6}{5}, \quad \varrho(2, 3) = \varrho(3, 2) = \frac{8}{5},$$

$$\varrho(3, 1) = \varrho(1, 3) = \frac{151}{100}.$$

and

$$\varsigma(1, 1) = \varsigma(2, 2) = \varsigma(3, 3) = 1, \quad \varsigma(1, 2) = \varsigma(2, 1) = \frac{6}{5}, \quad \varsigma(2, 3) = \varsigma(3, 2) = \frac{33}{20},$$

$$\varsigma(3, 1) = \varsigma(1, 3) = \frac{8}{3}.$$

Now define the self-mapping Υ on \mathbb{X} as follows:

$$\Upsilon 1 = \Upsilon 2 = \Upsilon 3 = 2$$

Now we will verify the condition 1:

Case 1: When $p = q = 1$,

$$|h(\Upsilon p, \Upsilon q)| = |h(\Upsilon(1), \Upsilon(1))| = |h(2, 2)| = 0 \leq l|h(1, 1)|.$$

Case 2: When $p = 1, q = 2$,

$$|h(\Upsilon p, \Upsilon q)| = |h(\Upsilon(1), \Upsilon(2))| = |h(2, 2)| = 0 \leq l|h(1, 2)|.$$

Case 3: When $p = 1, q = 3$,

$$|h(\Upsilon p, \Upsilon q)| = |h(\Upsilon(1), \Upsilon(3))| = |h(2, 2)| = 0 \leq l|h(1, 3)|.$$

Case 4: When $p = 2, q = 1$,

$$|h(\Upsilon p, \Upsilon q)| = |h(\Upsilon(2), \Upsilon(1))| = |h(2, 2)| = 0 \lesssim l|h(2, 1)|.$$

Case 5: When $p = q = 2$,

$$|h(\Upsilon p, \Upsilon q)| = |h(\Upsilon(2), \Upsilon(2))| = |h(2, 2)| = 0 \lesssim l|h(2, 2)|.$$

Case 6: When $p = 2, q = 3$,

$$|h(\Upsilon p, \Upsilon q)| = |h(\Upsilon(2), \Upsilon(3))| = |h(2, 2)| = 0 \lesssim l|h(2, 3)|.$$

Case 7: When $p = 3, q = 1$,

$$|h(\Upsilon p, \Upsilon q)| = |h(\Upsilon(3), \Upsilon(1))| = |h(2, 2)| = 0 \lesssim l|h(3, 1)|.$$

Case 8: When $p = 3, q = 2$,

$$|h(\Upsilon p, \Upsilon q)| = |h(\Upsilon(3), \Upsilon(2))| = |h(2, 2)| = 0 \lesssim l|h(3, 2)|.$$

Case 9: When $p = q = 3$,

$$|h(\Upsilon p, \Upsilon q)| = |h(\Upsilon(3), \Upsilon(3))| = |h(2, 2)| = 0 \lesssim l|h(3, 3)|.$$

For all $k \in (0, 1)$, it is clear that the above conditions are satisfied, these conditions are also satisfied for $\Upsilon 1 = \Upsilon 2 = \Upsilon 3 = 1$. For any $p_0 \in \mathbb{X}$ condition (2) holds along with conditions of Theorem 1.1. Therefore, there exists a unique fixed point at 1.

Given $p_0 \in \mathbb{X}$, the orbit $\mathbb{O}(u_0)$ of p_0 is defined as $\mathbb{O}(u_0) = \{p_0, \Upsilon p_0, \Upsilon^2 p_0, \dots\}$, where Υ is a self-map on the set \mathbb{X} . The operator $\Gamma : \mathbb{X} \rightarrow \mathbb{R}$ is called Υ -orbitally lower semi-continuous at $\varpi \in \mathbb{X}$ if when $\{p_n\}$ in $\mathbb{O}(p_0)$ such that $\lim_{n \rightarrow \infty} h_{cdl}(p_n, \varpi) = 0$, we get that $\Gamma(\varpi) \leq \liminf_{n \rightarrow \infty} \Gamma(p_n)$.

Proceeding similarly as [24] and using Definition 3, we have the following corollary generalizing the Theorem 1.1 in [13].

Let Υ be a self-map on (\mathbb{X}, h_{cdl}) a complete complex-valued double controlled metric-like space by two mappings ϱ, ς . Given $p_0 \in \mathbb{X}$. Let $l \in (0, 1)$ be such that

$$h_{cdl}(\Upsilon z, \Upsilon^2 z) \lesssim l h_{cdl}(z, \Upsilon z), \text{ for each } z \in \mathbb{O}(p_0).$$

Take $p_n = \Upsilon^n p_0$ and suppose that

$$\sup_{m \geq 1} \lim_{i \rightarrow \infty} \frac{\varrho(p_{i+1}, p_{i+2})}{\varrho(p_i, p_{i+1})} \varsigma(p_{i+1}, p_m) < \frac{1}{l}.$$

Then, $\lim_{n \rightarrow \infty} h_{cdl}(p_n, \kappa) = 0$. We also we have that $\Upsilon \kappa = \kappa$ if and only if the operator $x \mapsto h_{cdl}(x, \Upsilon x)$ is Υ -orbitally lower semi-continuous at p .

Our next fixed point result involve a Reich type inequality, as follows.

Theorem 1.2 Let (\mathbb{X}, h_{cdl}) be a CDCMLS by the functions $\varrho, \varsigma : \mathbb{X} \times \mathbb{X} \rightarrow [1, \infty)$ and Υ be a self mapping satisfying Reich condition. That is, Υx satisfies

$$h_{cdl}(\Upsilon p, \Upsilon q) \lesssim \alpha h_{cdl}(p, q) + \beta(p, \Upsilon p) + \gamma(q, \Upsilon q) \quad (8)$$

for $\alpha, \beta, \gamma \in (0, 1)$ with $\alpha + \beta + \gamma < 1$ and $\gamma = \frac{\alpha + \beta}{1 - \gamma} < 1$, for all $p, q \in \mathbb{X}$.

For $p_0 \in \mathbb{X}$ we choose $p_n = \Upsilon^n p_0$. Assume that

$$\sup_{m \geq 1} \lim_{l \rightarrow \infty} \frac{\varrho(p_{l+1}, p_{l+2})}{\varrho(p_l, p_{l+1})} \varsigma(p_{l+1}, p_m) < \frac{1}{l}. \quad (9)$$

$$\lim_{n \rightarrow \infty} \varrho(p, p_n) < \infty \text{ exist and finite and } \lim_{n \rightarrow \infty} \varsigma(p_n, p) < \frac{1}{\gamma} \quad (10)$$

Then, Υ has a unique fixed point.

Proof: Let $p_0 \in \mathbb{X}$. Consider the sequence $\{p_n\}$ with $p_{n+1} = \Upsilon p_n$ for all $n \in \mathbb{N}$. It is clear that if there exist n_0 for which $p_{n_0+1} = p_{n_0}$ then $\Upsilon p_{n_0} = p_{n_0}$. Then the proof is finished.

Thus, we suppose that $p_{n_0+1} \neq p_{n_0}$ for every $n \in \mathbb{N}$. Therefore, we may assume that $p_{n+1} \neq p_n$ for all $n \in \mathbb{N}$. Now

$$\begin{aligned} h_{cdl}(p_n, p_{n+1}) &= h_{cdl}(\Upsilon p_{n-1}, \Upsilon p_n) \lesssim \alpha h_{cdl}(p_{n-1}, p_n) + \beta h_{cdl}(p_{n-1}, \Upsilon p_{n-1}) + \gamma h_{cdl}(p_n, \Upsilon p_n) \\ &= \alpha h_{cdl}(p_{n-1}, p_n) + \beta h_{cdl}(p_{n-1}, \Upsilon p_n) + \gamma h_{cdl}(\Upsilon p_n, p_{n+1}). \end{aligned} \quad (11)$$

Therefore, we get

$$h_{cdl}(p_n, p_{n+1}) \lesssim \left(\frac{\alpha + \beta}{1 - \gamma} \right) h_{cdl}(p_{n-1}, p_n) = l h_{cdl}(p_{n-1}, p_n) \quad (12)$$

Thus, we obtain

$$h_{cdl}(p_n, p_{n+1}) \lesssim l h_{cdl}(p_{n-1}, p_n) \lesssim l^2 h_{cdl}(p_{n-2}, p_{n-1}) \lesssim \dots \lesssim l^m h_{cdl}(p_0, p_1). \quad (13)$$

for all $n, m \in \mathbb{N}$ with $n < m$

$$\begin{aligned} h_{cdl}(p_n, p_m) &\lesssim \varrho(p_n, p_{n+1}) h_{cdl}(p_n, p_{n+1}) + \varsigma(p_{n+1}, p_m) h_{cdl}(p_{n+1}, p_m) \\ &\lesssim \varrho(p_n, p_{n+1}) h_{cdl}(p_n, p_{n+1}) + \varsigma(p_{n+1}, p_m) \varrho(p_{n+1}, p_{n+2}) h_{cdl}(p_{n+1}, p_{n+2}) \\ &\quad + \varsigma(p_{n+1}, p_m) \varsigma(p_{n+2}, p_m) h_{cdl}(p_{n+2}, p_m) \\ &\lesssim \varrho(p_n, p_{n+1}) h_{cdl}(p_n, p_{n+1}) + \varsigma(p_{n+1}, p_m) \varrho(p_{n+1}, p_{n+2}) h_{cdl}(p_{n+1}, p_{n+2}) \\ &\quad + \varsigma(p_{n+1}, p_m) \varsigma(p_{n+2}, p_m) \varrho(p_n, p_{n+3}) h_{cdl}(p_{n+3}, p_m) \\ &\lesssim \dots \\ &\lesssim \varrho(p_n, p_{n+1}) h_{cdl}(p_n, p_{n+1}) + \sum_{i=n+1}^{m-2} \left(\prod_{j=n+1}^i \varsigma(p_j, p_m) \right) \varrho(p_i, p_{i+1}) h_{cdl}(p_i, p_{i+1}) \\ &\quad + \prod_{k=n+1}^{m-1} \varsigma(p_k, p_m) h_{cdl}(p_{m-1}, p_m) \\ &\lesssim \varrho(p_n, p_{n+1}) l^m h_{cdl}(p_0, p_1) + \sum_{i=n+1}^{m-2} \left(\prod_{j=n+1}^i \varsigma(p_j, p_m) \right) \varrho(p_i, p_{i+1}) l^i h_{cdl}(p_0, p_1) \\ &\quad + \prod_{i=n+1}^{m-1} \varsigma(p_i, p_m) l^{m-1} h_{cdl}(p_0, p_1) \end{aligned}$$

$$\begin{aligned}
 & \lesssim \mathfrak{Q}(p_n, p_{n+1}) l^m h_{cdl}(p_0, p_1) + \sum_{i=n+1}^{m-2} \left(\prod_{j=n+1}^i \varsigma(p_j, p_m) \right) \mathfrak{Q}(p_i, p_{i+1}) l^i h_{cdl}(p_0, p_1) \\
 & + \left(\prod_{i=n+1}^{m-1} \varsigma(p_i, p_m) \right) l^{m-1} \mathfrak{Q}(p_{m-1}, p_m) h_{cdl}(p_0, p_1) \\
 & = \mathfrak{Q}(p_n, p_{n+1}) l^m h_{cdl}(p_0, p_1) + \sum_{i=n+1}^{m-1} \left(\prod_{j=n+1}^i \varsigma(p_j, p_m) \right) \mathfrak{Q}(p_i, p_{i+1}) l^i h_{cdl}(p_0, p_1) \\
 & \lesssim \mathfrak{Q}(p_n, p_{n+1}) l^m h_{cdl}(p_0, p_1) + \sum_{i=n+1}^{m-1} \left(\prod_{j=0}^i \varsigma(p_j, p_m) \right) \mathfrak{Q}(p_i, p_{i+1}) l^i h_{cdl}(p_0, p_1). \\
 R_n &= \sum_{i=0}^n \left(\prod_{j=0}^i \varsigma(p_j, p_m) \right) \mathfrak{Q}(p_i, p_{i+1}) l^i h_{cdl}(p_1, p_0).
 \end{aligned}$$

Then we applying the ratio test, we have

$$\begin{aligned}
 g_n &= \left(\prod_{j=0}^i \varsigma(p_j, p_m) \right) \mathfrak{Q}(p_i, p_{i+1}) l^i h_{cdl}(p_1, p_0) \\
 \frac{g_{n+1}}{g_n} &= l \mathfrak{Q}(p_{i+1}, p_m) \frac{\varsigma(p_{i+1}, p_{i+2})}{\varsigma(p_i, p_{i+1})}
 \end{aligned}$$

Therefore under condition (9), the series $\sum_n g_n$ converges. Therefore, $\lim_{n \rightarrow \infty} R_n$ exist. So the real number sequence $\{R_n\}$ is Cauchy.

Thus we obtained the inequality

$$h_{cdl}(p_n, p_m) \lesssim h_{cdl}(p_1, p_0) [l^n \varsigma(p_n, p_{n+1}) + (R_{m-1} - R_n)].$$

Letting $n, m \rightarrow \infty$, we get

$$\lim_{n, m \rightarrow \infty} h_{cdl}(p_n, p_m) = 0,$$

so the sequence $\{p_n\}$ is Cauchy. Since (\mathbb{X}, h_{cdl}) is a complete CDCMLS, then there exists some $p_0^* \in \mathbb{X}$ such that

$$\lim_{n \rightarrow \infty} h_{cdl}(p_n, p_0^*) = 0$$

Which means $p_n \rightarrow p_0^*$ and $n \rightarrow \infty$.

Now, our claim is to show that $\Upsilon p_0^* = p_0^*$.

$$\begin{aligned}
 h_{cdl}(p_0^*, \Upsilon p_0^*) & \lesssim \mathfrak{Q}(p_0^*, p_{n+1}) h_{cdl}(p_0^*, p_{n+1}) + \varsigma(p_{n+1}, \Upsilon p_0^*) h_{cdl}(p_n, p_{n+1}) h_{cdl}(p_{n+1}, \Upsilon p_0^*) \\
 & = \mathfrak{Q}(p_0^*, p_{n+1}) h_{cdl}(p_0^*, p_{n+1}) + \varsigma(p_{n+1}, \Upsilon p_0^*) h_{cdl}(p_n, p_{n+1}) h_{cdl}(p_{n+1}, \Upsilon p_0^*) \\
 & \lesssim \mathfrak{Q}(p_0^*, p_{n+1}) h_{cdl}(p_0^*, p_{n+1}) + \varsigma(p_{n+1}, \Upsilon p_0^*) [\alpha h_{cdl}(p_n, \Upsilon p_0^*) + \beta h_{cdl}(p_n, \Upsilon p_n) + \gamma h_{cdl}(p_0^*, \Upsilon p_0^*)] \\
 & = \varsigma(p_0^*, p_{n+1}) h_{cdl}(p_0^*, p_{n+1}) + \mathfrak{Q}(p_{n+1}, \Upsilon p_0^*) [\alpha h_{cdl}(p_n, \Upsilon p_0^*) + \beta h_{cdl}(p_n, p_{n+1}) + \gamma h_{cdl}(p_0^*, \Upsilon p_0^*)]
 \end{aligned}$$

Using this facts in (10) and letting the limit as $n \rightarrow \infty$ we obtain

$$h_{cdl}(p_0^*, \Upsilon p_0^*) \lesssim \mathfrak{Q}(p_{n+1}, \Upsilon p_0^*) \left[\gamma \lim_{n \rightarrow \infty} h_{cdl}(p_0^*, \Upsilon(p_0, \Upsilon p_0^*)) \right].$$

Suppose that $\Upsilon p_0^* \neq p_0^*$. Since $\lim_{n \rightarrow \infty} \mathfrak{Q}(p_{n+1}, \Upsilon p_n) < \frac{1}{l}$ we have

$$0 < h_{cdl}(p_0^*, \Upsilon p_0^*) \lesssim \mathfrak{Q}(p_{n+1}, \Upsilon p_0^*) [\gamma h_{cdl}(p_0^*, \Upsilon p_0^*)] \\ \lesssim \mathfrak{Q}(p_0^*, \Upsilon p_0^*).$$

It is a contradiction, which means $p_0^* = \Upsilon p_0^*$.

Finally, assume that Υ has two fixed points, say p and q .

Then $h_{cdl}(p, q) = h_{cdl}(\Upsilon p, \Upsilon q) \lesssim \alpha h_{cdl}(p, q) + \beta h_{cdl}(p, \Upsilon p) + \gamma h_{cdl}(q, \Upsilon q)$ and so $h_{cdl}(p, q)(1 - \alpha) \lesssim 0$. Since $\alpha \neq 1$. We get $h_{cdl}(p, q) = 0$ which implies $p = q$.

This completes the proof.

Let $\mathbb{E} = \{1, 2, 3\}$. Define $h = h_{cdl} : \mathbb{E} \times \mathbb{E} \rightarrow \mathbb{C}$ by

$$h(1, 1) = h(2, 2) = 0 \text{ and } h(3, 3) = \frac{i}{2}, \\ h(1, 2) = h(2, 1) = 2 + i, \quad h(2, 3) = h(3, 2) = i, \\ h(1, 3) = h(3, 1) = 1 - i.$$

Define $\mathfrak{Q}, \varsigma : \mathbb{E} \times \mathbb{E} \rightarrow [1, \infty)$ by

$$\mathfrak{Q}(1, 1) = \mathfrak{Q}(2, 2) = \mathfrak{Q}(3, 3) = 1, \\ \mathfrak{Q}(1, 2) = \mathfrak{Q}(2, 1) = 1, \quad \mathfrak{Q}(2, 3) = \mathfrak{Q}(3, 2) = \frac{8}{7}, \quad \mathfrak{Q}(3, 1) = \mathfrak{Q}(1, 3) = \frac{3}{2}.$$

and

$$\varsigma(1, 1) = \varsigma(2, 2) = \varsigma(3, 3) = 1, \\ \varsigma(1, 2) = \varsigma(2, 1) = \frac{7}{6}, \quad \varsigma(2, 3) = \varsigma(3, 2) = \frac{9}{2}, \quad \varsigma(3, 1) = \varsigma(1, 3) = 1.$$

Let

$$\Upsilon(1) = 2, \Upsilon(2) = 2, \Upsilon(3) = 2$$

Proof: $h(\Upsilon p, \Upsilon q) \lesssim \alpha h(p, q) + \beta h(p, \Upsilon p) + \gamma h(q, \Upsilon q)$.

Case 1: When $p = 1, q = 2$

$$|h(\Upsilon p, \Upsilon q)| = |h(2, 2)| = 0 \lesssim \frac{7\sqrt{5}}{12} = \frac{1}{3}\sqrt{5} = \alpha|h(1, 2)| + \beta|h(1, 2)| + \gamma|h(2, 2)| \\ = \alpha|h(p, q)| + \beta|h(p, \Upsilon p)| + \gamma|h(q, \Upsilon q)|.$$

Case 2: When $p = 1, q = 1$

$$|h(\Upsilon p, \Upsilon q)| = |h(2, 2)| = 0 \lesssim \frac{13\sqrt{5}}{36} = \frac{1}{3}\sqrt{5} + \frac{1}{9}\sqrt{5} = \alpha|h(1, 1)| + \beta|h(1, 2)| + \gamma|h(1, 2)| \\ = \alpha|h(p, q)| + \beta|h(p, \Upsilon p)| + \gamma|h(q, \Upsilon q)|.$$

Case 3: When $p = 2, q = 2$

$$|h(\Upsilon p, \Upsilon q)| = |h(2, 2)| = 0 \lesssim 0 = 0 + 0 + 0 = \alpha|h(2, 2)| + \beta|h(2, 2)| + \gamma|h(2, 2)| \\ = \alpha|h(p, q)| + \beta|h(p, \Upsilon p)| + \gamma|h(q, \Upsilon q)|.$$

Case 4: When $p = 3, q = 3$

$$|h(\Upsilon p, \Upsilon q)| = |h(2, 2)| = 0 \lesssim \frac{12 + 13\sqrt{2}}{36\sqrt{2}} = \frac{1}{3\sqrt{2}} + \frac{1}{4} + \frac{1}{9} = \alpha|h(3, 3)| + \beta|h(3, 2)| + \gamma|h(3, 2)| \\ = \alpha|h(p, q)| + \beta|h(p, \Upsilon p)| + \gamma|h(q, \Upsilon q)|.$$

Case 5: When $p = 2, q = 1$

$$\begin{aligned} |h(\Upsilon p, \Upsilon q)| &= |h(2, 2)| = 0 \lesssim \frac{4\sqrt{5}}{9} = \frac{1}{3}\sqrt{5} + 0 + \frac{1}{9}\sqrt{5} = \alpha|h(2, 1)| + \beta|h(2, 2)| + \gamma|h(1, 2)| \\ &= \alpha|h(p, q)| + \beta|h(p, \Upsilon p)| + \gamma|h(q, \Upsilon q)|. \end{aligned}$$

Case 6: When $p = 3, q = 1$

$$\begin{aligned} |h(\Upsilon p, \Upsilon q)| &= |h(2, 2)| = 0 \lesssim \frac{12\sqrt{2} + 9 + 4\sqrt{5}}{36} = \frac{1}{3}\sqrt{2} + \frac{1}{4} + \frac{1}{9}\sqrt{5} \\ &= \alpha|h(3, 1)| + \beta|h(3, 2)| + \gamma|h(1, 2)| \\ &= \alpha|h(p, q)| + \beta|h(p, \Upsilon p)| + \gamma|h(q, \Upsilon q)|. \end{aligned}$$

Case 7: When $p = 1, q = 3$

$$\begin{aligned} |h(\Upsilon p, \Upsilon q)| &= |h(2, 2)| = 0 \lesssim \frac{12\sqrt{2} + 9\sqrt{5} + 4}{36} = \frac{1}{3}\sqrt{2} + \frac{1}{4}\sqrt{5} + \frac{1}{9} \\ &= \alpha|h(1, 3)| + \beta|h(1, 2)| + \gamma|h(3, 2)| \\ &= \alpha|h(p, q)| + \beta|h(p, \Upsilon p)| + \gamma|h(q, \Upsilon q)|. \end{aligned}$$

Case 8: When $p = 2, q = 3$

$$\begin{aligned} |h(\Upsilon p, \Upsilon q)| &= |h(2, 2)| = 0 \lesssim \frac{4}{9} = \frac{1}{3} + \frac{1}{4}(0) + \frac{1}{9} = \alpha|h(2, 3)| + \beta|h(2, 2)| + \gamma|h(3, 2)| \\ &= \alpha|h(p, q)| + \beta|h(p, \Upsilon p)| + \gamma|h(q, \Upsilon q)|. \end{aligned}$$

For all $\alpha, \beta, \gamma \in (0, 1)$ with $\alpha + \beta + \gamma < 1$, it is clear that the above conditions are satisfied, these conditions are also satisfied for $\Upsilon 1 = \Upsilon 2 = \Upsilon 3 = 2$. For any $p_0 \in \mathbb{E}$ condition (9) holds along with conditions of theorem 1.2. Therefore, there exists a unique fixed point at 2.

4. Existence and uniqueness of the solution of a Fredholm type integral equation

During this section we suppose the following Fredholm integral equation

$$p(u) = f(u) + \int_a^b B(u, v, p(v))dv, \quad u, v \in [a, b], p(u) \in \mathbb{X}, \quad (14)$$

where $B(u, v, p(v)) : [a, b] \times [a, b] \times \mathbb{C} \rightarrow \mathbb{C}$ and $f(u) : [a, b] \rightarrow \mathbb{C}$ be two bounded and continuous functions.

To prove the existence of solution for integral eq. (14) we use Theorem 1.1. Then we give the following result.

Theorem 1.3 Let $\mathbb{X} = C([a, b], \mathbb{C})$ is the set of all continuous and complex valued functions which are defined on $[a, b]$. Also let $\Upsilon : \mathbb{X} \rightarrow \mathbb{X}$ be an operator defined as:

$$p(u) = f(u) + \int_a^b B(u, v, p(v))dv, \quad u, v \in [a, b]. \quad (15)$$

Suppose the following conditions hold:

- i. the functions $B(u, v, p(v)) : [a, b] \times [a, b] \times \mathbb{C} \rightarrow \mathbb{C}$ and $f(u) : [a, b] \rightarrow \mathbb{C}$ it's a continuous function;
- ii. $|B(u, v, p(v)) - B(u, v, q(v))| \lesssim \frac{1}{\tau\sqrt{b-a}} |p(u) - q(u)|$, for all $p, q \in \mathbb{X}$ and $\omega \in (1, \frac{1}{\lambda}]$ with $\lambda \in (0, 1)$.

Then Eq. (14) has a unique solution.

Proof: Let $\mathbb{X} = C([a, b], \mathbb{C})$ and $h_{cdl} : \mathbb{X} \times \mathbb{X} \rightarrow \mathbb{C}$ such that,

$$h_{cdl}(p, q) = \|p - q\|_{\infty} = |p(u) - q(u)|^2 e^{i \cos^{-1} \tau},$$

where $|x| = \sqrt{\alpha^2 + \beta^2}$, with $\alpha, \beta \in \mathbb{R}$, $\tau > 0$ and $i = \sqrt{-1} \in \mathbb{C}$.

Let $\vartheta_u, \varrho_u : \mathbb{X} \times \mathbb{X} \rightarrow [1, \infty)$ be defined as

$$\varrho_u(p, q) = \begin{cases} 1, & \text{if } p, q \in [0, 1], \\ \max\{p(u), q(u)\}, & \text{otherwise.} \end{cases}$$

$$\varsigma_u(p, q) = \begin{cases} 1, & \text{if } p, q \in [0, 1], \\ \frac{1 + \max\{p(u), q(u)\}}{\min\{p(u), q(u)\}}, & \text{otherwise.} \end{cases}$$

We observe that (\mathbb{X}, h_{cdl}) is a complete CDCMLS. Then the problem (14) can be translated to find a fixed point of the operator Υ .

Then we have the next inequality

$$\begin{aligned} |\Upsilon p(u) - \Upsilon q(u)|^2 &\lesssim \left| \int_a^b B(u, v, p(v)) dv - \int_a^b B(u, v, q(v)) dv \right|^2 \\ &\lesssim \int_a^b |B(u, v, p(v)) - B(u, v, q(v))|^2 dv \\ &\lesssim \frac{1}{\tau^2(b-a)} \int_a^b |p(v) - q(v)|^2 dv \\ &= \frac{e^{-i \cos^{-1} \tau}}{\tau^2(b-a)} \int_a^b |p(v) - q(v)|^2 e^{i \cos^{-1} \tau} dv \\ &= \frac{e^{-i \cos^{-1} \tau}}{\tau^2(b-a)} \|p - q\|_{\infty} \left(\int_a^b dv \right). \end{aligned}$$

Following the calculus we obtain

$$|\Upsilon p(u) - \Upsilon q(u)|^2 e^{i \cos^{-1} \tau} = \|\Upsilon p - \Upsilon q\|_{\infty} \lesssim \frac{1}{\tau^2} |p(u) - q(u)|^2 e^{i \cos^{-1} \tau} = \frac{1}{\tau^2} \|p - q\|_{\infty}.$$

Using the hypothesis (ii) we have

$$h_{cdl}(\Upsilon p, \Upsilon q) = \|\Upsilon p - \Upsilon q\|_{\infty} \lesssim \frac{1}{\tau^2} \|p - q\|_{\infty} = \frac{1}{\tau^2} h_{cdl}(p, q).$$

It is easy to check that, for both cases of the expressions of $\varrho_u(p, q)$ and $\varsigma_u(p, q)$, the conditions (2) and (3) are true.

Then, for $0 < \delta = \frac{1}{\tau^2} < 1$, all the hypothesis of Theorem 1.1 holds. In this conditions, we get that Eq. (14) has a unique solution.

5. Conclusions

Continuing in the same way as in Refs. [12, 25], we have introduced the concept of complex-valued double controlled metric-like spaces(CDCMLS). Some fixed point results and supporting examples in this setting, the related Banach contraction principle, the Reich type fixed point results, are presented. Moreover, the last section is dedicated to apply our main result in order to prove the existence and uniqueness of the solution of a Fredholm type integral equation.

Conflict of interest

The authors declare no conflict of interest.

Author details

Mohammad S.R. Chowdhury^{1*}, Muhammad Suhail Aslam¹,
Musawa Yahya Almusawa² and Muhammad Imran Asjad³


1 Faculty of Science, Department of Mathematics and Statistics, University of Lahore, Lahore, Pakistan

2 Faculty of Science, Department of Mathematics, Jazan University, Jazan, Saudi Arabia

3 Department of Mathematics, University of Management and Technology, Lahore, Pakistan

*Address all correspondence to: showkat.rahim@math.uol.edu.pk

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Banach S. Sur les opérations dans les ensembles abstraits et leur applications aux équations intégrales. *Fundamenta Mathematicae*. 1922;3(1):133-181
- [2] Bakhtin IA. The contraction mapping principle in almost metric spaces. *Journal of Functional Analysis*. 1989;30:26-37
- [3] Czerwik S. Contraction mappings in b -metric spaces. *Acta Mathematica et Informatica Universitatis Ostraviensis*. 1993;1:5-11
- [4] Aslam MS, Chowdhury MSR, Guran L, Alqudah MA, Abdeljawad T. Fixed point theory in complex valued controlled metric spaces with an application. *AIMS Mathematics*. 2022;7(7):11879-11904
- [5] Kamran T, Samreen M, Ul. Ain Q. A generalization of b -metric space and some fixed point theorems. *Mathematics*. 2017;5:1-7
- [6] Mlaiki N, Aydi H, Souayah N, Abdeljawad T. Controlled metric type spaces and the related contraction principle. *Mathematics*. 2018;6:194
- [7] Abdeljawad T, Mlaiki N, Aydi H, Souayah N. Double controlled metric type spaces and some fixed point results. *Mathematics*. 2018;6:320. DOI: 10.3390/math6120320
- [8] Panda SK, Abdeljawad T, Ravichandran C. A complex valued approach to the solutions of Riemann-Liouville integral, Atangana-Baleanu integral operator and non-linear telegraph equation via fixed point method. *Chaos, Solitons and Fractals*, Elsevier. 2020;2020:130
- [9] Ullah N, Shagari MS, Azam A. Fixed point theorems in complex valued extended b -metric spaces. *Moroccan Journal of Pure and Applied Analysis*. 2019;5(2):140-163
- [10] Abdeljawad T, Abodayeh K, Mlaiki N. On fixed point generalizations to partial b -metric spaces. *Journal of Computational Analysis and Applications*. 2015;19:883-891
- [11] Rao KP, Swamy P, Prasad J. A common fixed point theorem in complex valued b -metric spaces. *Bulletin of Mathematics and Statistics Research*. 2013;1(1):2013
- [12] Mlaiki N. Double controlled metric-like spaces. *Journal of Inequalities and Applications*. 2020;2020:189. DOI: 10.1186/s13660-020-02456-z
- [13] Hicks TL, Rhodes BE. A Banach type fixed point theorem. *Mathematical Society of Japan*. 1979;24:327-330
- [14] Matkowski J. Fixed point theorems for mappings with a contractive iterate at a point. *Proceedings of American Mathematical Society*. 1977;62:344-348
- [15] Kannan R. Some results on fixed points. *Bulletin of the Calcutta Mathematical Society*. 1968;60:71-76
- [16] Afshari H, Atapour M, Aydi H. Generalized $\varphi - \psi$ - Geraghty multivalued mappings on b -metric spaces endowed with a graph. *Journal of Applied and Engineering Mathematics*. 2017;7:248-260
- [17] Aslam MS, Bota MF, Chowdhury MSR, Guran L, Saleem N. Common fixed points technique for existence of a solution of Urysohn type integral equations system in complex valued b -metric spaces. *Mathematics*. 2021;9:400. DOI: 10.3390/math9040400

[18] Alharbi N, Aydi H, Felhi A, Ozel C, Sahmim S. ϱ -Contractive mappings on rectangular b -metric spaces and an application to integral equations. *Journal of Mathematical Analysis and Applications*. 2018;**9**:47-60

[19] Aydi H, Karapinar E, Bota MF, Mitrović S. A fixed point theorem for set-valued quasi-contractions in b -metric spaces. *Journal of Fixed Point Theory and Applications*. 2012;**88**:2012

[20] Aydi H, Bota MF, Karapinar E, Moradi S. A common fixed point for weak ϕ -contractions on b -metric spaces. *Fixed Point Theory*. 2012;**13**:337-346

[21] Harandi AA. Metric-like spaces, partial metric spaces and fixed points. *Fixed Point Theory and Applications*. 2012;**2012**:204

[22] Azam A, Fisher B, Khan M. Common fixed point theorems in complex valued metric spaces. *Numerical Functional Analysis and Optimization*. 2011;**32**(3): 243-253

[23] Hosseini A, Karizaki MM. On the complex valued metric-like spaces. *arXiv*. 2022;**2022**(1):1-8

[24] Kamran T, Samreen M, QUL A. A generalization of b -metric space and some fixed point theorems. *Mathematics*. 2017;**5**:19

[25] Aysegul T. On double controlled metric-like spaces and related fixed point theorems. *Advances in the Theory of Nonlinear Analysis and its Applications*. 2021;**2**:167-172

Section 2

Geometry and Topology

Geometry of Sub-Riemannian Manifolds Equipped with a Quasi-Semi-Weyl Structure

Sergey Galaev and Evgeny Kokin

Abstract

Quasi-semi-Weyl and quasi-statistical structures are based on a connection with torsion. In this chapter, as a connection with torsion, we consider the so-called extended connection, which is defined with the help of an intrinsic connection, i.e., a connection in the distribution of a sub-Riemannian manifold, as well as with the help of an endomorphism that preserves the indicated distribution and is called a structural endomorphism. It is proved that the extended connection is a connection of the quasi-semi-Weyl structure of a sub-Riemannian manifold of contact type only if the distribution of the sub-Riemannian manifold is involutive. In order to be able to consider sub-Riemannian manifolds with a not necessarily involutive distribution, the concepts of sub-Riemannian quasi-semi-Weyl and sub-Riemannian quasi-statistical structures are introduced, which are modifications of quasi-semi-Weyl and quasi-statistical structures for the case of sub-Riemannian manifolds of contact type. The structural endomorphism for the connection of a sub-Riemannian quasi-statistical structure is found. As an example, we consider non-holonomic Kenmotsu manifolds, which are sub-Riemannian manifolds of contact type endowed with an additional structure. It is proved that the restriction of the structural endomorphism to the distribution of such manifold differs from the identity transformation only by a factor.

Keywords: sub-Riemannian manifold of contact type, intrinsic geometry of a sub-Riemannian manifold, extended connection, sub-Riemannian quasi-statistical structure, sub-Riemannian quasi-semi-Weyl structure

1. Introduction

A torsion-free linear connection ∇ on a manifold M is said to be metric compatible with g if the metric tensor is a *Codacci tensor* with respect to ∇ :

$$\nabla_X g(Y, Z) = \nabla_Y g(X, Z), \quad X, Y, Z \in \Gamma(TM). \quad (1)$$

The statistical structure introduced by Lauritzen in [1] is a pair (g, ∇) , where g is a pseudo-Riemannian metric and ∇ is a torsion-free linear connection compatible with

it. A manifold M equipped with a statistical structure is called a statistical manifold. Statistical structures are related to the theory of conjugate linear connections developed by A.P. Norden [2]. The set of torsion-free connections conjugate with respect to the metric constitutes a very interesting class of connections compatible with the metric. This class, along with the Levi-Civita connection, which is the only self-adjoint connection, includes other connections that are of interest to researchers. Linear connections compatible with the Riemannian metric find interesting applications in the geometric interpretation of a number of problems in mathematical statistics [3–5].

To describe geometric structures in the spaces of quantum states, it is convenient to use quasi-statistical manifolds, assuming the presence of torsion in a connection compatible with the metric. In [6], the concept of a quasi-semi-Weyl structure was introduced and methods for constructing quasi-statistical and quasi-semi-Weyl structures were proposed.

In this chapter, analogues of the quasi-statistical and quasi-semi-Weyl structures are introduced, these are the sub-Riemannian quasi-statistical structure and the sub-Riemannian quasi-semi-Weyl structure, respectively. The refinement of previously known concepts is caused by the desire to take into account the specifics of sub-Riemannian manifolds of contact type, on which these structures are specified. A sub-Riemannian manifold of contact type is a smooth manifold M of dimension n with a sub-Riemannian structure $(M, \vec{\xi}, \eta, g, D)$ given on it, where η is a 1-form generating the distribution D , $D = \ker(\eta)$, $\vec{\xi}$ is a vector field that generates a rigging D^\perp of the distribution D , $D^\perp = \text{span}(\vec{\xi})$, g is a Riemannian metric on the manifold M with respect to which the distributions D and D^\perp are mutually orthogonal. In this case, the equalities $\eta(\vec{\xi}) = 1$, $g(\vec{\xi}, \vec{\xi}) = 1$, $d\eta(\vec{\xi}, \cdot) = 0$ hold true.

A sub-Riemannian quasi-statistical manifold is a sub-Riemannian manifold of contact type endowed with a sub-Riemannian quasi-statistical structure (M, g, ∇^N) . Here ∇^N is a connection with torsion of a special type, which is called in this work an extended connection, or an N -connection. An N -connection ∇^N is defined by a pair (∇, N) , where ∇ is an intrinsic linear connection [7], $N : TM \rightarrow TM$ is an endomorphism of the tangent bundle of the manifold M such that $N\vec{\xi} = \vec{0}$, $N(D) \subset D$. We will say that the N -connection ∇^N is the extension of a connection ∇ by means of the endomorphism N .

An extended connection ∇^N is defined as the only connection that satisfies the following conditions:

(1) $\nabla_X^N Y \in \Gamma(D)$, (2) $\nabla_X^N \vec{\xi} = \vec{0}$, (3) $\nabla_{\vec{\xi}}^N Y = [\vec{\xi}, Y] + NY$, (4) $\nabla_Y^N Z = \nabla_Y Z$, $X, Y, Z \in \Gamma(TM)$; $Y, Z \in \Gamma(D)$. If $\nabla_X g(Y, Z) = 0$, $\nabla_X Y - \nabla_Y X - P[X, Y] = 0$, where $X, Y, Z \in \Gamma(D)$ and $P : D \oplus D^\perp \rightarrow D$ is a projector, then ∇ is called an intrinsic metric connection. The torsion $T(X, Y)$ of an intrinsic linear connection is given by the equality $T(X, Y) = \nabla_X Y - \nabla_Y X - P[X, Y]$.

It can be directly verified that the torsion $S(X, Y)$ of the connection ∇^N is given by the following formula [7, 8]:

$$S(X, Y) = 2\omega(X, Y)\vec{\xi} + \eta(X)NY - \eta(Y)NX, \quad X, Y, Z \in \Gamma(TM). \quad (2)$$

Here $\omega(X, Y) = d\eta(X, Y)$.

The interest to the connections with torsion is due to their use in theoretical physics [9–11]. An N -connection ∇^N was defined by one of the authors of this chapter and was studied by him, for example, in [7, 12–15]. By appropriately specifying the endomorphism N , one can obtain most of the previously known classes of connections with torsion, e.g., the Schouten-van Kampen connection, the Tanaka-Webster connection, etc. [7, 16]. At the same time, in those papers in which connections with torsion ∇^N were used (for specific endomorphisms N), the presence of an endomorphism N was not explicitly discussed. An exception is the work [17] (see also [13]), where the properties of the endomorphism N , denoted in the paper by the symbol “ τ ”, received a special attention.

Motivation for defining and studying sub-Riemannian quasi-statistical structures (M, g, ∇^N) is supported by the following facts:

1. As mentioned above, the class of extended connections ∇^N is widely represented in modern geometric studies;
2. N -connections ∇^N arise naturally as continuations of the intrinsic connections ∇ , which occupy an important place in the geometric modeling of problems in non-holonomic mechanics and theoretical physics [7].

In the present work, we define a sub-Riemannian quasi-statistical structure as a triple (M, g, ∇^N) , where M is a sub-Riemannian manifold, and the connection ∇^N is related to the metric g through the equality

$$\nabla_X^N g(Y, Z) - \nabla_Y^N g(X, Z) + \tilde{S}(X, Y, Z) - 2\omega(X, Y)\eta(Z) = 0, \quad (3)$$

where $\tilde{S}(X, Y, Z) = g(S(X, Y), Z)$, $X, Y, Z \in \Gamma(TM)$, $\omega = d\eta$.

The extension of the term “quasi-statistical structure” by adding the word “sub-Riemannian” is justified by the use in the structure-defining equality of the external form $\omega = d\eta$, referring to the intrinsic geometry of a sub-Riemannian manifold of contact type [7]. Although a sub-Riemannian quasi-statistical structure is defined on a sub-Riemannian manifold of contact type, as a basis manifold we will often use almost contact metric manifolds as sub-Riemannian manifolds carrying an additional structure.

2. Main results

An almost contact metric manifold is a smooth manifold M of odd dimension $n = 2m + 1$, $m \geq 1$, with an almost contact metric structure $(M, \vec{\xi}, \eta, \varphi, g)$ defined on it [8, 16]. Here, in particular, η is a 1-form and $\vec{\xi}$ is a vector field generating, respectively, the distribution $D: D = \ker(\eta)$ and the rigging D^\perp of the distribution $D: D^\perp = \text{span}(\vec{\xi})$. A smooth distribution D is called the distribution of an almost contact metric manifold. There is the decomposition $TM = D \oplus D^\perp$. An almost contact metric manifold is called normal if the condition $N_\varphi + 2d\eta \otimes \vec{\xi} = 0$ is satisfied, where

$N_\varphi(X, Y) = [\varphi X, \varphi Y] + \varphi^2[X, Y] - \varphi[\varphi X, Y] - \varphi[X, \varphi Y]$ is the Nijenhuis tensor of the endomorphism φ .

To carry out the necessary calculations, it is convenient to use the adapted coordinates [8]. A chart $k(x^i)(i, j, k = 1, \dots, n; a, b, c = 1, \dots, n-1)$ on the manifold M is said to be adapted to the distribution D if $\partial_n = \vec{\xi}$ [7]. Let $P : TM \rightarrow D$ be the projection defined by the decomposition $TM = D \oplus D^\perp$ and $k(x^a)$ be an adapted chart. The vector fields $P(\partial_a) = \vec{e}_a = \partial_a - \Gamma_a^n \partial_n$ generate the distribution $D : D = \text{Span}(\vec{e}_a)$. For a non-holonomic frame field $(\vec{e}_i) = (\vec{e}_a, \partial_n)$, the relation $[\vec{e}_a, \vec{e}_b] = 2\omega_{ba}\partial_n$ holds true. We additionally require the condition $\omega(\vec{\xi}, \cdot) = 0$, which is equivalent to the condition $\partial_n \Gamma_a^n = 0$. The condition $\omega(\vec{\xi}, \cdot) = 0$ does not impose strong restrictions on the structures under consideration. For example, all normal almost contact metric manifolds satisfy this condition.

Let $k(x^i)$ and $k'(x^{i'})$ be adapted charts, then we obtain the following coordinate transformation formulas:

$$x^a = x^a(x^{a'}), \quad x^n = x^{n'} + x^n(x^{a'}). \quad (4)$$

An intrinsic linear connection ∇ [8] on an almost contact metric manifold is a map

$$\nabla : \Gamma(D) \times \Gamma(D) \rightarrow \Gamma(D) \quad (5)$$

satisfying the following conditions:

1. $\nabla_{f_1 X + f_2 Y} = f_1 \nabla_X + f_2 \nabla_Y$,
2. $\nabla_X f Y = (Xf)Y + f \nabla_X Y$,
3. $\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z$,

where $\Gamma(D)$ is the module of admissible vector fields (vector fields belonging to the distribution D at each point of the manifold).

Let $\tilde{\nabla}$ be the Levi-Civita connection and $\tilde{\Gamma}_{jk}^i$ be its Christoffel symbols. Using equality

$$2\Gamma_{ij}^m = g^{km}(e_i g_{jk} + e_j g_{ik} - e_k g_{ij} + \Omega_{kj}^l g_{li} + \Omega_{ki}^l g_{lj}) + \Omega_{ij}^m, \quad (6)$$

where $[\vec{e}_i, \vec{e}_j] = \Omega_{ij}^m \vec{e}_k$, we prove the following proposition [8].

Proposition 1. The Christoffel symbols $\tilde{\Gamma}_{ij}^k$ of the Levi-Civita connection of a sub-Riemannian manifold with respect to the adapted coordinates have the following form:

$$\begin{aligned} \tilde{\Gamma}_{ab}^c &= \Gamma_{ab}^c, \tilde{\Gamma}_{ab}^n = \omega_{ba} - C_{ab}, \tilde{\Gamma}_{an}^b = \tilde{\Gamma}_{na}^b = C_a^b + \psi_a^b, \tilde{\Gamma}_{na}^n = -\partial_n \Gamma_a^n, \tilde{\Gamma}_{nn}^a = g^{ab} \partial_n \Gamma_b^n, \\ \text{where } \Gamma_{bc}^a &= \frac{1}{2} g^{ad} (\vec{e}_b g_{cd} + \vec{e}_c g_{bd} - \vec{e}_d g_{bc}), \psi_a^b = g^{bc} \omega_{ac}, C_{ab} = \frac{1}{2} \partial_n g_{ab}, C_a^b = g^{bc} C_{ac}. \end{aligned} \quad (7)$$

Here the endomorphism $\psi : TM \rightarrow TM$ is determined by the equality $\omega(X, Y) = g(\psi X, Y)$. The following relations also hold: $C(X, Y) = \frac{1}{2} (L_{\vec{\xi}} g)(X, Y)$, $g(CX, Y) = C(X, Y)$.

Note that in the case when $\omega\left(\vec{\xi}, \cdot\right) = 0$, the expressions with respect to the adapted coordinates for the Christoffel symbols $\tilde{\Gamma}_{ij}^k$ of the Levi-Civita connection of the sub-Riemannian manifold take a simpler form. The following Christoffel symbols remain non-zero:

$$\tilde{\Gamma}_{ab}^c = \Gamma_{ab}^c, \tilde{\Gamma}_{ab}^n = \omega_{ba} - C_{ab}, \tilde{\Gamma}_{an}^b = \tilde{\Gamma}_{na}^b = C_a^b + \psi_a^b. \quad (8)$$

Let ∇ be an intrinsic linear connection, let ∇^N be a connection uniquely determined by the conditions (1) $\nabla_X^N Y \in \Gamma(D)$, (2) $\nabla_X^N \vec{\xi} = \vec{0}$, (3) $\nabla_{\vec{\xi}}^N Y = \left[\vec{\xi}, Y\right] + NY$, (4) $\nabla_Y^N Z = \nabla_Y Z$, $X \in \Gamma(TM)$; $Y, Z \in \Gamma(D)$. Throughout the paper, we will assume that the torsion $T(X, Y)$ of the intrinsic linear connection ∇ is equal to zero:

$$T(X, Y) = \nabla_X Y - \nabla_Y X - P[X, Y] = 0.$$

It can be checked directly that the torsion $S(X, Y)$ of the extended connection ∇^N may be found by the following formula [15]:

$$S(X, Y) = 2\omega(X, Y)\vec{\xi} + \eta(X)NY - \eta(Y)NX, X, Y, Z \in \Gamma(TM). \quad (9)$$

Here

$$\omega(X, Y) = d\eta(X, Y).$$

If $\nabla_X g(Y, Z) = 0$, $X, Y, Z \in \Gamma(D)$, then the following equality holds true:

$$\nabla_X^N Y = \tilde{\nabla}_X Y + (\tilde{\nabla}_X \eta)(Y)\vec{\xi} - \eta(Y)\tilde{\nabla}_X \vec{\xi} - \eta(X)(C + \psi - N)Y. \quad (10)$$

It can be directly verified that in this case the non-zero Christoffel symbols G_{jk}^i of the connection ∇^N with respect to the adapted coordinates have the form

$$G_{bc}^a = \frac{1}{2}g^{ad}\left(\vec{e}_b g_{cd} + \vec{e}_c g_{bd} - \vec{e}_d g_{bc}\right), G_{na}^b = N_a^b. \quad (11)$$

In the case of intrinsic linear connection (not necessarily metric), the coefficients G_{bc}^a may be found from the relation $\nabla_a \vec{e}_b = G_{ab}^c \vec{e}_c$.

Let $\tilde{S}(X, Y, Z) = g(S(X, Y), Z)$, $X, Y, Z \in \Gamma(TM)$. With respect to the adapted coordinates, the non-zero components of the tensor $\tilde{S}(X, Y, Z) = g(S(X, Y), Z)$, $X, Y, Z \in \Gamma(TM)$, have the following form:

$$\begin{aligned} \tilde{S}(\vec{e}_a, \vec{e}_b, \partial_n) &= 2\omega_{ab}, \\ \tilde{S}(\vec{e}_a, \partial_n, \vec{e}_b) &= -g(N\vec{e}_a, \vec{e}_b), \\ \tilde{S}(\partial_n, \vec{e}_a, \vec{e}_b) &= g(N\vec{e}_a, \vec{e}_b). \end{aligned}$$

The concept of a quasi-statistical structure on a Riemannian manifold is discussed in [1]. A triplet (M, g, ∇) is called a quasi-statistical structure if the following equality holds [1]:

$$\nabla_X g(Y, Z) - \nabla_Y g(X, Z) + \tilde{S}(X, Y, Z) = 0, \quad X, Y, Z \in \Gamma(TM). \quad (12)$$

We will call the triplet (M, g, ∇^N) a sub-Riemannian quasi-statistical structure (SRCS-structure) if the equality

$$\Phi(X, Y, Z) = \nabla_X^N g(Y, Z) - \nabla_Y^N g(X, Z) + \tilde{S}(X, Y, Z) - 2\omega(X, Y)\eta(Z) = 0,$$

where $\tilde{S}(X, Y, Z) = g(S(X, Y), Z)$, $X, Y, Z \in \Gamma(TM)$, $\omega = d\eta$, holds true. In this case, we will assume that the torsion of the intrinsic connection ∇ is equal to zero. Note that if (M, g, ∇^N) is a sub-Riemannian quasi-statistical structure, then

$$\nabla_X g(Y, Z) - \nabla_Y g(X, Z) = 0, \quad X, Y, Z \in \Gamma(D). \quad (13)$$

We claim that if the triplet (M, g, ∇^N) is a sub-Riemannian quasi-statistical structure, then the corresponding intrinsic connection ∇ is compatible with the metric g restricted to the distribution D .

Theorem 1. A triple (M, g, ∇^N) is an SRCS-structure of a sub-Riemannian manifold M of contact type if and only if $N = 2C$ and

$$\nabla_X g(Y, Z) - \nabla_Y g(X, Z) = 0, \quad X, Y, Z \in \Gamma(D). \quad (14)$$

Proof.

Let us consider the equality

$$\nabla_X^N g(Y, Z) - \nabla_Y^N g(X, Z) + \tilde{S}(X, Y, Z) - 2\omega(X, Y)\eta(Z) = 0, \quad X, Y, Z \in \Gamma(TM) \quad (15)$$

We fix adapted coordinates. If $X = \partial_n, Y = \vec{e}_b, Z = \vec{e}_c$, then we get

$$\nabla_n^N g(\vec{e}_b, \vec{e}_c) + g(N\vec{e}_b, \vec{e}_c) = 0.$$

Hence it holds

$$\partial_n g_{bc} - N_{bs}^d g_{da} - N_{cs}^d g_{bd} + N_{bs}^d g_{da} = 0.$$

Finally, we get the equality

$$N_b^a = 2C_b^a. \quad (16)$$

This proves the theorem.

Consider, as an example, a non-holonomic Kenmotsu manifold. The Kenmotsu manifolds were discovered in 1972 in [18]. The concept of a non-holonomic Kenmotsu manifold was introduced in [19]. In [20], a non-holonomic Kenmotsu manifold is equipped with an N -connection ∇^N with torsion of a special type. A normal almost contact metric manifold M is called a non-holonomic Kenmotsu manifold if $L_{\xi}g = 2(g - \eta \otimes \eta)$ holds true [19].

Let now (M, g, ∇^N) be an SRCS-structure. Turning to the equation

$$\nabla_X^N g(Y, Z) - \nabla_Y^N g(X, Z) + \tilde{S}(X, Y, Z) - 2\omega(X, Y)\eta(Z) = 0, \quad (17)$$

we obtain the following modification of Theorem 1.

Theorem 2. If ∇ is an intrinsic metric connection, then the triple (M, g, ∇^N) is a sub-Riemannian quasi-statistical structure on a non-holonomic Kenmotsu manifold M if and only if $N = 2\tilde{E}$, where $\tilde{E}X = X, \tilde{E}\vec{\xi} = \vec{0}, X \in \Gamma(D)$.

Thus with respect to adapted coordinates, the non-zero Christoffel symbols G_{jk}^i of the connection ∇^N in the case under consideration have the form

$$G_{bc}^a = \frac{1}{2}g^{ad}(\vec{e}_b g_{cd} + \vec{e}_c g_{bd} - \vec{e}_d g_{bc}), G_{na}^b = 2\delta_b^a. \quad (18)$$

Let $\tilde{\nabla}^N$ be the connection conjugate to the connection ∇^N included in the SRCS-structure (M, g, ∇^N) :

$$Xg(Y, Z) = g(\nabla_X^N Y, Z) + g(Y, \tilde{\nabla}_X^N Z). \quad (19)$$

Denote by \tilde{G}_{jk}^i the Christoffel symbols of the connection $\tilde{\nabla}^N$ with respect to adapted coordinates. For $X = \vec{e}_a, Y = \vec{e}_b, Z = \vec{e}_c$, we get

$$\tilde{G}_{bc}^a = G_{bc}^a. \quad (20)$$

In the case when $X = \partial_n, Y = \vec{e}_b, Z = \vec{e}_c$, we have

$$\tilde{G}_{na}^b = 0. \quad (21)$$

All other Christoffel symbols \tilde{G}_{jk}^i are zero.

Proposition 2. The non-zero Christoffel symbols \tilde{G}_{jk}^i of the connection $\tilde{\nabla}^N$ have the form: $\tilde{G}_{bc}^a = G_{bc}^a$. In the case of a metric connection ∇ it holds

$$\tilde{G}_{bc}^a = \frac{1}{2}g^{ad}(\vec{e}_b g_{cd} + \vec{e}_c g_{bd} - \vec{e}_d g_{bc}). \quad (22)$$

Thus, it is shown that the conjugate connection to a connection ∇^N also has the structure of an N-connection with zero structural endomorphism N .

The following statement is an analogue of Proposition 1 from [6].

Proposition 3. Let Q be a tensor field of type $(1, 2)$ such that $g(Q(X, Z), Y) = g(X, Q(Y, Z))$. Then (M, g, ∇^N) is a sub-Riemannian quasi-statistical structure if and only if $(M, g, \dot{\nabla}^N = \nabla^N + Q)$ is a sub-Riemannian quasi-statistical structure.

Proof. Let us use the equalities obtained in [6],

$$\dot{S}(X, Y) = S(X, Y) + Q(X, Y) - Q(Y, X), \quad (23)$$

$$\dot{\nabla}_X^N g(Y, Z) = \nabla_X^N g(Y, Z) - g(Q(X, Y), Z) - g(Y, Q(X, Z)). \quad (24)$$

This implies

$$\dot{\Phi}(X, Y, Z) = \Phi(X, Y, Z) + g(X, Q(Y, Z)) - g(Y, Q(X, Z)), X, Y, Z \in \Gamma(TM). \quad (25)$$

Here

$$\dot{\Phi}(X, Y, Z) = \dot{\nabla}_X^N g(Y, Z) - \dot{\nabla}_Y^N g(X, Z) + \dot{S}(X, Y, Z) - 2\omega(X, Y)\eta(Z) = 0. \quad (26)$$

Thus the proposition is proved.

Below is the definition of a quasi-semi-Weyl structure adapted to our case, first published in [6]. A triple (M, g, ∇^N) is called a quasi-semi-Weyl structure if the following equality holds:

$$\nabla_X^N g(Y, Z) + \eta(X)g(Y, Z) = \nabla_Y^N g(X, Z) + \eta(Y)g(X, Z) - \tilde{S}(X, Y, Z). \quad (27)$$

We rewrite the last equality with respect to adapted coordinates:

1. If $X = \vec{e}_a, Y = \vec{e}_b, Z = \partial_n$, then it follows that $d\eta = 0$;
2. If $X = \partial_n, Y = \vec{e}_b, Z = \vec{e}_c$, then we get:

$$\nabla_n^N g(\vec{e}_b, \vec{e}_c) + g(\vec{e}_b, \vec{e}_c) = -g(N\vec{e}_b, \vec{e}_c).$$

Hence with respect to adapted coordinates we have

$$\partial_n g_{bc} - N_{bs}^d g_{da} - N_{cs}^d g_{bd} + g_{bc} = -N_{bs}^d g_{da}. \quad (28)$$

Finally, we get the equality

$$N_b^a = 2C_b^a + \delta_b^a. \quad (29)$$

Thus the following theorem holds true.

Theorem 3. A triple (M, g, ∇^N) is a quasi-semi-Weyl structure on a sub-Riemannian M manifold of contact type if and only if the following conditions are satisfied:

1. The distribution D of the manifold M is involutive;
2. Structural endomorphism N is of the following form:

$$N = 2C + \tilde{E}, \text{ where, } \tilde{E}X = X, \tilde{E}\vec{\xi} = \vec{0}, X \in \Gamma(D);$$

3. The intrinsic connection ∇ is compatible with the metric g restricted to the distribution D

$$\nabla_X g(Y, Z) - \nabla_Y g(X, Z) = 0, X, Y, Z \in \Gamma(D). \quad (30)$$

In order to abandon the requirement that the distribution D of the manifold M is involutive, we refine the definition of a quasi-semi-Weyl structure for the case of a sub-Riemannian manifold of contact type. Namely, let us call the triple (M, g, ∇^N) a sub-Riemannian quasi-semi-Weyl structure if the following equality holds

$$\nabla_X^N g(Y, Z) + \eta(X)g(Y, Z) = \nabla_Y^N g(X, Z) + \eta(Y)g(X, Z) - \tilde{S}(X, Y, Z) + 2\omega(X, Y)\eta(Z), \quad X, Y, Z \in \Gamma(TM). \quad (31)$$

Let us rewrite the last equality with respect to adapted coordinates.

If $X = \partial_n, Y = \vec{e}_b, Z = \vec{e}_c$, then we get:

$$\nabla_n^N g(\vec{e}_b, \vec{e}_c) + g(\vec{e}_b, \vec{e}_c) = -g(N\vec{e}_b, \vec{e}_c). \quad (32)$$

Hence with respect to adapted coordinates it holds that

$$\partial_n g_{bc} - N_b^d g_{da} - N_c^d g_{bd} + g_{bc} = -N_b^d g_{da}. \quad (33)$$

Finally, we get the equality

$$N_b^a = 2C_b^a + \delta_b^a. \quad (34)$$

Thus the following theorem holds true.

Theorem 4. A triple (M, g, ∇^N) is a sub-Riemannian quasi-semi-Weyl structure of a sub-Riemannian manifold M of contact type if and only if

$$1. N = 2C + \tilde{E}, \text{ where } \tilde{E}X = X, \quad \tilde{E}\vec{\xi} = \vec{0}, \quad X \in \Gamma(D); \quad (35)$$

$$2. \nabla_X g(Y, Z) - \nabla_Y g(X, Z) = 0, \quad X, Y, Z \in \Gamma(D). \quad (36)$$

The following theorem establishes connections between the concepts of a sub-Riemannian quasi-statistical structure and a sub-Riemannian quasi-semi-Weyl structure of a sub-Riemannian manifold M of contact type. It is a consequence of Theorems 1 and 4.

Theorem 5. Let the connections ∇^N and $\nabla^{\tilde{N}}$ be extensions of the same intrinsic connection ∇ on a sub-Riemannian manifold M of contact type. Then the triple (M, g, ∇^N) is a sub-Riemannian quasi-statistical structure if and only if the triple $(M, g, \nabla^{\tilde{N}})$ is a sub-Riemannian quasi-semi-Weyl structure with

$$\dot{N} = N + \tilde{E} \quad (37)$$

3. Conclusion


This chapter makes an additional contribution to the theory of N-connections used for the development and numerous applications of contact-type sub-Riemannian geometry. The connection ∇^N , which is an extension of the intrinsic connection ∇ through endomorphism N , in a certain sense organically corresponds to the ideas of geometrization of a vast class of problems in theoretical physics. Above have spoken about the contribution of the theory of N-connections to the development of the geometry of Einstein manifolds and their generalizations [7]. The role of N-connections in the well-known geometric interpretation of the motion of a charged particle in the unified theory of gravitational and electromagnetic interactions has yet to be essentially clarified [9, 21].

Author details

Sergey Galaev and Evgeny Kokin*
Saratov State University, Saratov, Russia

*Address all correspondence to: evgeny@epromicro.com

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Lauritzen SL. Statistical manifolds. In: *Differential Geometry in Statistical Inferences; IMS Lecture Notes Monograph Series*. Vol. 10. Hayward, CA, USA: Institute of Mathematical Statistics; 1987. pp. 96-163
- [2] Norden AP. *Spaces of Affine Connection*. 2nd ed. Moscow: Nauka; 1976. p. 432
- [3] Morozova EA, Chentsov NN. Natural geometry of families of probabilistic laws. *Itogi nauki i tekhniki. VINITI. Modern Problems Mathematics Fundamental Directions*. 1991;**83**: 133-265
- [4] Rylov AA. Connections compatible with the Riemannian metric in the theory of statistical manifolds. *Russian Mathematics (Izvestiya VUZ Matematika)* 1994;**3**:62-64
- [5] Stepanov SE, Stepanova ES, Shandra IG. Conjugate connections on statistical manifolds. *Izvestiya VUZ Matematika*. 2007;**10**:90-98
- [6] Blaga AM, Nannicini A. On statistical and Semi-Weyl manifolds admitting Torsion. *Mathematics*. 2022;**10**:990. DOI: 10.3390/math10060990
- [7] Galaev SV. ∇^N -Einstein almost contact metric manifolds, *Tomsk State University Bulletin. Mathematics and Mechanics*. 2021;**70**:5-15
- [8] Galaev SV. Intrinsic geometry of almost contact Kählerian manifolds. *Acta Mathematica Academiae Paedagogicae Nyiregyhaziensis*. 2015;**31**(1):35-46
- [9] Crimea VR. The Jacobi equation for horizontal geodesics on a nonholonomic distribution and the Schouten curvature tensor. *Differential Equations and Control Processes*. 2018;**3**:64-94 Available from: <http://diffjournal.spbu.ru/RU/numbers/2018.3/article.1.3.html>
- [10] Agricola I, Ferreira AC. Einstein manifolds with skew torsion. *The Quarterly Journal of Mathematics*. 2014; **65**:717-741
- [11] Friedrich T, Ivanov S. Parallel spinors and connections with skew-symmetric torsion in string theory. *Asian Journal of Mathematics*. 2002;**6**:303-336
- [12] Galaev SV. The golden section in the geometry of η -Einstein Sub-Riemannian manifolds with N -connection, science statements of the Belgorod State University. *Series: Mathematics Physics*. 2019;**51**(4):S. 465-S. 474. DOI: 10.18413/2075-4639-2019-51-4-465-474
- [13] Galaev SV. Geometric interpretation of the Wagner curvature tensor for the case of a manifold with a contact metric structure. *Siberian Mathematical Journal*. 2016;**57**(3(337)):632-640
- [14] Galaev SV. Almost contact metric spaces with N -connectivity. *Izvestiya of Saratov University. Mathematics, Mechanics and Informatics*. 2015;**15**(3):258-263
- [15] Galaev SV. Generalized Wagner curvature tensor of almost contact metric spaces. *Chebyshevskii sbornik*. 2016;**17**(3(59)):53-63
- [16] Bejancu A. Kahler contact distributions. *Journal of Geometry and Physics*. 2010;**60**:1958-1967
- [17] Falbel E, Gorodski C. On contact subriemannian symmetric spaces. *Annales Scientifiques de l'École Normale Supérieure, Serie 4*. 1995;**28**(5): 571-589

[18] Kenmotsu K. A class of almost contact Riemannian manifolds. *Tohoku Mathematical Journal*. 1972;**24**:93-103

[19] Bukusheva AV. On the geometry of non-holonomic Kenmotsu manifolds. *Izv. Altai State University*. 2021;**1**(117): 84-87

[20] Bukusheva AV. Non-holonomic Kenmotsu manifolds equipped with a generalized Tanaka–Webster connection. *Differential Geometry of Manifolds*. 2021;**52**:42-51

[21] Krym VR. Equations of geodesic for a charged particle in the unified theory of gravitational and electromagnetic interactions. *Teoreticheskaya i Matematicheskaya Fizika*. 1999;**119**(3): 811-820

Obtaining the Main Curvatures of Orientable Hypersurfaces, Based on Differential Geometry

Emmanoel Ferreira

Abstract

Most of the literature on differential geometry presents the coefficients of the first and second fundamental form to simplify the calculation of the curvatures on a surface and also to obtain other important information, such as the area of a surface. In this text, as the interest lies on the generalization of the idea of surface (hypersurface), such simplification by the use of these coefficients was not possible, in view of the complexity of the mathematical operations involved in the calculation of curvatures (if n = number of variables in the implicit function of the surface > 3), opting to use the linear operator $-DN_p$. To facilitate the calculation of the normal vector, the surface was described as the graph of a differentiable function $f: \mathbb{R}^{n-1} \rightarrow \mathbb{R}$. The example of the sphere thoroughly illustrates the procedure of calculating the main curvatures of a surface in the space \mathbb{R}^3 , and such procedure is extensive to the space \mathbb{R}^n . Some examples from the engineering area in which the variables of the implicit functions of the hypersurfaces are random were solved, and the results of the main curvatures, calculated by the proposed procedure, were exact and coincident with those provided in the literature.

Keywords: differential geometry, main curvatures, parametrization of a regular surface, orientable surface, implicit function

1. Introduction

This work focuses on providing the principal curvatures of hypersurfaces, from the normal vector, considering that much of the literature on the subject provides the results of these curvatures for the space \mathbb{R}^3 and allows for the calculation of the principal curvatures in space \mathbb{R}^n via the linear operator $-DN_p$. This procedure will be exemplified analytically for a surface in \mathbb{R}^3 (sphere). Furthermore, examples of hypersurfaces used in second-order reliability analysis (e.g., see [1]) of engineering problems with main curvatures results provided by MATLAB software (e.g., see [2]) will be presented.

The applications of the second-order reliability method (SORM), in engineering problems (e.g., see [1, 3–11]), in recent years, suggest a relevant interest in the SORM, and thus, there is much room for new research in this area, the main challenge being

to calculate the main curvatures of the limit state surface, which involves a lot of mathematical complexity and computational effort.

The proposed procedure extends to the space R^n , providing exact values for the main curvatures of any orientable hypersurface with deterministic variables, as in mathematics, or random variables, and in reliability analysis problems in engineering.

2. Procedure to obtain the main curvatures by differential geometry

In this section, the procedure is established, by differential geometry, to calculate the main curvatures of the surface at the p point.

2.1 Surfaces in R^3

To get a better understanding, in this subsection, the main curvatures at the p point of a surface in R^3 are calculated by differential geometry, and in the Subsection 2.2, the generalization of this process to the space R^n is performed.

2.1.1 Parameterization of a regular surface in R^3

The graph of an equation of the form $F(x,y,z) = 0$, where F is a differentiable function and its partial derivatives do not cancel each other, simultaneously, at any p point, such that $F(p) = 0$, is an example of a regular surface in R^3 . It is verified that the graph of a differentiable function $f: R^2 \rightarrow R$ is also an example of regular surface.

More generally, a subset S of R^3 is called regular surface if, for each p point $\in S$, there is an open neighborhood $V \subset R^3$ of p , an open $U \subset R^2$ and a bijection $\varphi: U \rightarrow V \cap S$, where φ is the form $\varphi(u,v) = \{x(u,v), y(u,v), z(u,v)\}$, with the properties described as follows (e.g., see [12]):

- a. φ is Class C^∞ , i.e., φ has continuous partial derivatives of all orders at the p point;
- b. φ is a homeomorphism (i.e., its inverse is continuous); and.
- c. for any point $q \in U$, the Jacobian matrix of φ has rank two. The referred matrix has rank two, which means that the image of the linear transformation obtained has dimension two, i.e., eliminating a line, conveniently chosen, the resulting 2×2 matrix has a determinant different from zero. The Jacobian matrix, in this case, has dimensions 3×2 , represented by.

$$J = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \\ \frac{\partial z}{\partial u} & \frac{\partial z}{\partial v} \end{bmatrix} \quad (1)$$

In these conditions, it is said that φ is a parameterization for S , as illustrated in **Figure 1**:

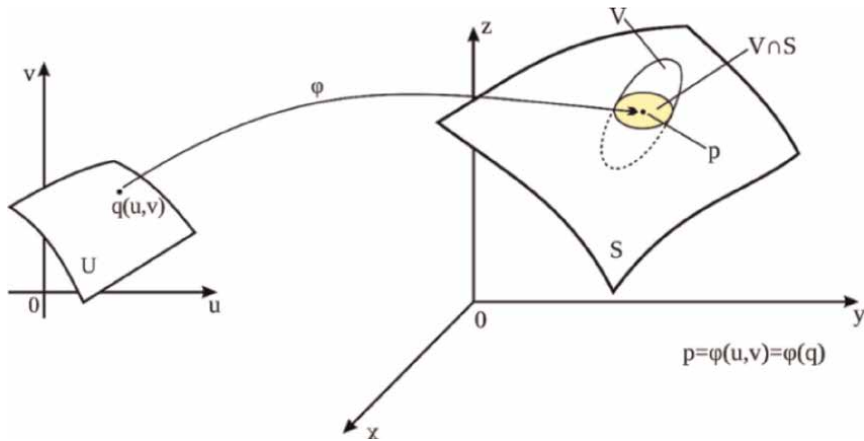


Figure 1.
 Parameterization of a regular surface. Source: Adapted from Ref. [13].

A regular surface $S \subset \mathbb{R}^3$ is orientable, if and only if there is a differentiable field $\mathbf{N}: S \rightarrow \mathbb{R}^3$ of normal vectors in S , according to Ref. [13].

2.1.2 Curvatures of a surface in \mathbb{R}^3

Being S an orientable surface, the Gauss application is the field of normal vectors $\mathbf{N}: S \rightarrow S^2$, where $S^2 \subset \mathbb{R}^3$ is the sphere of radius 1 and center at origin. \mathbf{N} is a differentiable application and its derivative $-DN_p: T_p S \rightarrow T_p S$ is an endomorphism (i.e., a linear transformation $T: U \rightarrow V$, being $U=V$), where $T_p S$ is the space (plane) tangent to S surface at the $p = \varphi(u, v)$ point. From the definition of derivative (rule of the chain), highlighted by Ref. [14], one has to.

$$-\mathbf{N}_u = -DN_{\varphi(u,v)}(\boldsymbol{\varphi}_u) \quad (2)$$

and

$$-\mathbf{N}_v = -DN_{\varphi(u,v)}(\boldsymbol{\varphi}_v)$$

where $\boldsymbol{\varphi}_u$ and $\boldsymbol{\varphi}_v$ are partial derivatives of parameterization $\varphi(u, v)$, i.e., they are tangent vectors that generate the plane $T_p S$.

The vectors \mathbf{N} and $\boldsymbol{\varphi}_u$ are orthogonal, as well as \mathbf{N} and $\boldsymbol{\varphi}_v$. Derivating the scalar products $\langle \boldsymbol{\varphi}_u, \mathbf{N} \rangle = 0$ and $\langle \boldsymbol{\varphi}_v, \mathbf{N} \rangle = 0$, it was concluded that $-DN_p$ is a self-adjunct linear application of $T_p S$ in $T_p S$. Thus, according to Ref. [14], the eigenvalues $k_1(p)$ and $k_2(p)$ of the linear operator $(-DN_p)$ are named main curvatures of S at the point p , and the orthogonal directions defined in $T_p S$ by eigenvalues $k_1(p)$ and $k_2(p)$ are named main directions.

2.1.3 Normal curvature

Being $\alpha: (a, b) \rightarrow S$ be a curve parameterized by arc length. The normal curvature of α in $\alpha(s)$ is the component of $\alpha''(s)$ according to the normal to S at this point and is given by $k_n(\alpha, s) = \langle \alpha''(s), \mathbf{N} \rangle$ (inner product between $\alpha''(s)$ and \mathbf{N} , with \mathbf{N} applied in the point $\alpha(s)$) and illustrated as **Figure 2**. If the curve was not

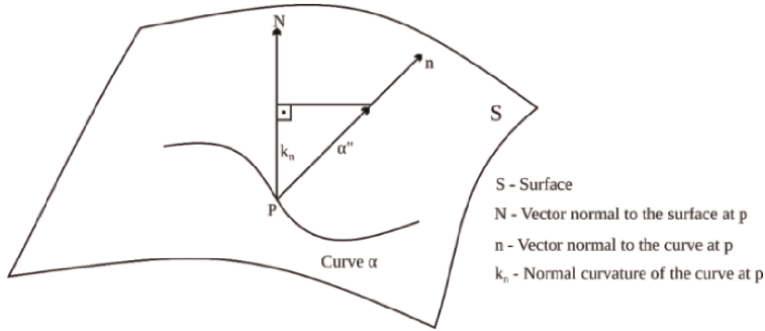


Figure 2.
 Normal curvature at point p. Source: The author.

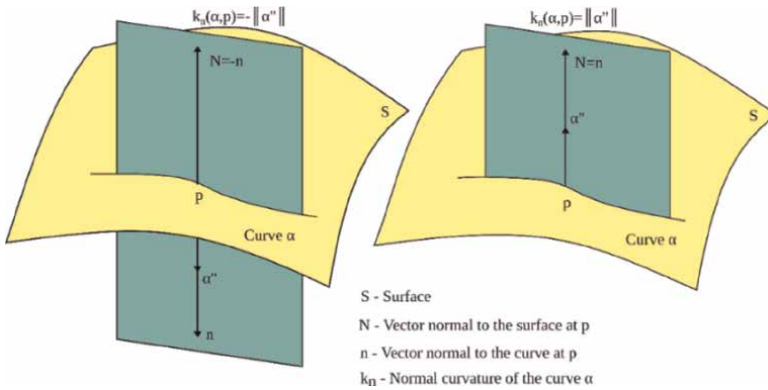


Figure 3.
 Normal curvatures when alpha is a normal section in point p. Source: The author.

parameterized by arc length, the formula of the normal curvature, according to Ref. [12], becomes.

$$k_n(\alpha, t) = \frac{1}{\|\alpha'(t)\|^2} \langle \alpha''(t), N \circ \alpha(t) \rangle \quad (3)$$

According to Ref. [12], the maximum and minimum values of the normal curvatures of the normal sections at p are the main curvatures of the surface in point p, as illustrated in **Figure 3**.

2.2 Surfaces in R^n

Most of the literature on differential geometry shows the coefficients of the first and second fundamental form to simplify the calculation of the curvatures in a surface in R^3 and also to obtain other information, such as a surface area. Herein, as the interest is the generalization of the surface idea (hypersurface), such simplification using these coefficients could not be done, due to the complexity of mathematical operations involved in calculating the curvatures when $n > 3$, and hence, in Subsection 2.1, it was opted to use the linear operator $-DN_p$.

To simplify calculation of normal vector to surface, it is described as the graph of a differentiable function $f: \mathbb{R}^{n-1} \rightarrow \mathbb{R}$.

2.2.1 Parameterization of the surface $g(V) = 0$

A parameterization for the surface in this vicinity can be given by.

$$\varphi_{(p)}(V_1, V_2, \dots, V_{n-1}) = [V_1, V_2, \dots, V_{n-1}, f(V_1, V_2, \dots, V_{n-1})] \quad (4)$$

The function $f(\mathbf{V})$, with $\mathbf{V} = (V_1, V_2, \dots, V_{n-1}) \in \mathbb{R}^{n-1}$, is obtained by explicitness for any variable of the vector \mathbf{V} of the function $g(\mathbf{V}) = 0$, where $\mathbf{V} = (V_1, V_2, \dots, V_{n-1}, V_n) \in \mathbb{R}^n$. Considering, for example, the explicitness of the last variable of $g(\mathbf{V}) = 0$; it has $V_n = f(V_1, V_2, \dots, V_{n-1})$.

2.2.2 Obtaining the vectors tangent to the surface $g(V) = 0$

The vectors tangent, which correspond to partial derivatives of Eq. (4), is calculated in point p according to.

$$\begin{aligned} \boldsymbol{\varphi}_{V_1}(p) &= (1, 0, \dots, 0, f_{V_1}(p)) = \left(1, 0, \dots, 0, \frac{\partial f(p)}{\partial V_1}\right) \\ \boldsymbol{\varphi}_{V_i}(p) &= (0, \dots, 1, \dots, 0, f_{V_i}(p)) = \left(0, \dots, 1, \dots, 0, \frac{\partial f(p)}{\partial V_i}\right); 1 < i < n-1 \\ \boldsymbol{\varphi}_{V_{n-1}}(p) &= (0, 0, \dots, 0, 1, f_{V_{n-1}}(p)) = \left(0, 0, \dots, 1, \frac{\partial f(p)}{\partial V_{n-1}}\right) \end{aligned} \quad (5)$$

2.2.3 Obtaining the normal vector to the surface $g(V) = 0$ and its partial derivatives

The normal vector at point p is calculated by extending the equation shown in Ref. [13] for this vector, i.e.,

$$\mathbf{N}(p) = \frac{(-f_{V_1}(p), -f_{V_2}(p), \dots, -f_{V_{n-1}}(p), 1)}{\sqrt{(f_{V_1}(p))^2 + (f_{V_2}(p))^2 + \dots + (f_{V_{n-1}}(p))^2 + 1}} \quad (6)$$

The partial derivatives of the normal vector are obtained by

$$\mathbf{N}_{V_j}(p) = \frac{\partial \mathbf{N}(p)}{\partial V_j}, j = 1, 2, \dots, n-1 \quad (7)$$

2.2.4 Obtaining the main curvatures of the surface $g(V) = 0$

Once performed the calculation of the normal vector and its partial derivatives \mathbf{N}_{V_j} , they can be written as a linear combination of the vectors, $\boldsymbol{\varphi}_{V_1}, \dots, \boldsymbol{\varphi}_{V_{n-1}}$, of the tangent plane, obtaining the matrix (\mathbf{M}) of the linear operator $-\mathbf{D}\mathbf{N}_p$, whose eigenvalues are the main curvatures. Extending the equation shown by Ref. [14] for the referred linear operator, it has

$$(-DN_p) \left(\Phi_{V_j} \right) = -N_{V_j}, j = 1, 2, \dots, n-1 \quad (8)$$

thus:

$$\begin{aligned} -N_{V_1} &= (-N_{1,1}) \cdot \Phi_{V_1} + \dots + (-N_{1,n-1}) \cdot \Phi_{V_{n-1}} \\ &\vdots \\ -N_{V_{n-1}} &= (-N_{n-1,1}) \cdot \Phi_{V_1} + \dots + (-N_{n-1,n-1}) \cdot \Phi_{V_{n-1}} \end{aligned} \quad (9)$$

and

$$M = \begin{bmatrix} -N_{1,1} & \cdot & \cdot & -N_{1,n-1} \\ -N_{2,1} & -N_{2,2} & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ -N_{n-1,1} & -N_{n-1,2} & \cdot & -N_{n-1,n-1} \end{bmatrix} \quad (10)$$

2.3 Example solved analytically by the proposed procedure

Compute the main curvatures, at point $p \left(0, \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2} \right)$, of a sphere (S) of radius equal to 1 and center at the origin, illustrated through **Figure 4**, whose equation is given by $g(V) = V_1^2 + V_2^2 + V_3^2 - 1 = 0$.

By explaining, for example, the last variable (V_3) of $g(V) = 0$ (it could have been chosen the variables V_1 or V_2) this sphere can be obtained through the function $f: \mathbb{R}^2 \rightarrow \mathbb{R}$ defined by

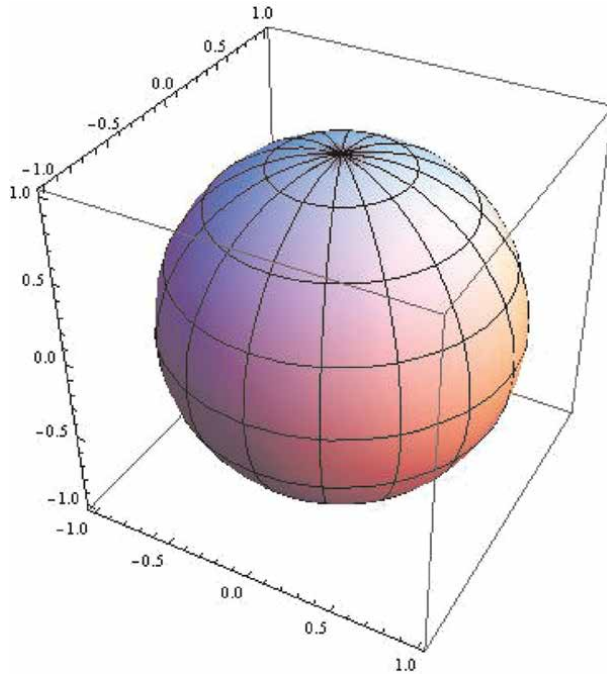


Figure 4.
Sphere (S) of radius equal to 1 and center at the origin. Source: The author.

$$V_3 = f(V_1, V_2) = \sqrt{1-V_1^2-V_2^2}$$

thus generating the parameterization

$$\varphi(V_1, V_2) = \left\{ V_1, V_2, \sqrt{1-V_1^2-V_2^2} \right\}$$

The partial derivatives of $f(V_1, V_2)$ are

$$f_{V_1}(V_1, V_2) = \frac{\partial f(V_1, V_2)}{\partial V_1} = -\frac{V_1}{\sqrt{1-V_1^2-V_2^2}}$$

and

$$f_{V_2}(V_1, V_2) = \frac{\partial f(V_1, V_2)}{\partial V_2} = -\frac{V_2}{\sqrt{1-V_1^2-V_2^2}}$$

According to Ref. [13]

$$N(V_1, V_2) = \frac{(-f_{V_1}, -f_{V_2}, 1)}{\sqrt{(f_{V_1})^2 + (f_{V_2})^2 + 1}} \quad (11)$$

defines the normal vectors to the surface.

Using Eq. (11), the result is for the normal vector

$$N(V_1, V_2) = \left\{ V_1, V_2, \sqrt{1-V_1^2-V_2^2} \right\}$$

Performing the partial derivatives of the normal vector in relation to V_1 and V_2 , we have

$$N_{V_1}(V_1, V_2) = \left\{ 1, 0, -\frac{V_1}{\sqrt{1-V_1^2-V_2^2}} \right\}$$

and

$$N_{V_2}(V_1, V_2) = \left\{ 0, 1, -\frac{V_2}{\sqrt{1-V_1^2-V_2^2}} \right\}$$

Considering the point $p\left(0, \frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}\right)$ that corresponds, in the parameterization $\varphi(V_1, V_2)$, to the point $q = \varphi\left(0, \frac{\sqrt{2}}{2}\right)$ and applying the equations obtained at that point, we have

$$N_{V_1}\left(0, \frac{\sqrt{2}}{2}\right) = (1, 0, 0)$$

and

$$\mathbf{N}_{v_2} \left(0, \frac{\sqrt{2}}{2} \right) = (0, 1, -1)$$

Following this procedure, the vectors that form the basis of the tangent plane $T_p S$ are determined. At point p , we have

$$\boldsymbol{\varphi}_{V_1}(V_1, V_2) = \left\{ 1, 0, -\frac{V_1}{\sqrt{1-V_1^2-V_2^2}} \right\}$$

and

$$\boldsymbol{\varphi}_{V_2}(V_1, V_2) = \left\{ 0, 1, -\frac{V_2}{\sqrt{1-V_1^2-V_2^2}} \right\}$$

because

$$\boldsymbol{\varphi}_{V_1}(V_1, V_2) = \frac{\partial(\boldsymbol{\varphi}(V_1, V_2))}{\partial V_1}$$

and

$$\boldsymbol{\varphi}_{V_2}(V_1, V_2) = \frac{\partial(\boldsymbol{\varphi}(V_1, V_2))}{\partial V_2}$$

At the point considered, we have $V_1 = 0$ e $V_2 = \frac{\sqrt{2}}{2}$, so

$$\boldsymbol{\varphi}_{V_1} \left(0, \frac{\sqrt{2}}{2} \right) = \{1, 0, 0\}$$

and

$$\boldsymbol{\varphi}_{V_2} \left(0, \frac{\sqrt{2}}{2} \right) = \{0, 1, -1\}$$

To determine the linear operator matrix $(-DN_p)$, one must apply this operator to the vectors $\boldsymbol{\varphi}_{V_1}$ and $\boldsymbol{\varphi}_{V_2}$ and write the results as a linear combination of $\boldsymbol{\varphi}_{V_1}$ e $\boldsymbol{\varphi}_{V_2}$

As reported by Ref. [14], one has that

$$(-DN_p) (\boldsymbol{\varphi}_{V_1}) = -\mathbf{N}_{v_1} \text{ and } (-DN_p) (\boldsymbol{\varphi}_{V_2}) = -\mathbf{N}_{v_2} \quad (12)$$

Thus, one can write $-\mathbf{N}_{v_1}$ and $-\mathbf{N}_{v_2}$ as a linear combination of the vectors $\boldsymbol{\varphi}_{V_1}$ and $\boldsymbol{\varphi}_{V_2}$. From the information obtained above, we have, at the point considered,

$$-\mathbf{N}_{v_1}(-1, 0, 0) = (-1) \boldsymbol{\varphi}_{V_1} + (0) \boldsymbol{\varphi}_{V_2}$$

and

$$-\mathbf{N}_{v_2}(0, -1, 1) = (0) \boldsymbol{\varphi}_{v_1} + (-1) \boldsymbol{\varphi}_{v_2}$$

Therefore, the matrix of the linear operator is

$$\mathbf{M} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

whose determinant, which is the Gaussian curvature, is $K = 1$ and whose eigenvalues, which are the main curvatures, are $k_1(p) = -1$ and $k_2(p) = -1$. The main curvatures values are exact according to the Ref. [15].

2.4 Examples solved by the proposed procedure via MATLAB software

Four examples were solved, being two from the mathematics area, aiming to validate the proposed procedure, and the last two from the engineering area.

2.4.1 The main curvature of a curve (parabola of degree 2)

Determine the main curvature of the parabola in **Figure 5** at the point (1,1).

The result of the main curvature is shown in **Table 1** and coincides with that presented in Ref. [12].

2.4.2 Main curvatures of a torus

Determine the main curvatures of the torus in **Figure 6** at point (7.1568, 7.1568, and 2.1214) with implicit function

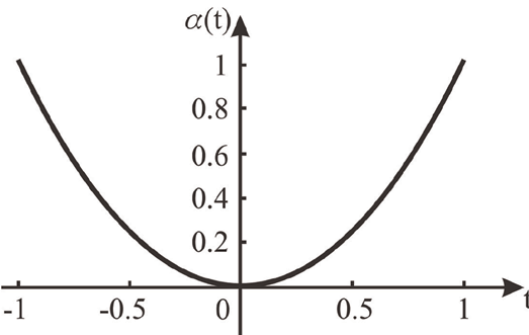


Figure 5.
 Graphical representation of a plane curve—2nd degree parabola. Source: The author.

Implicit function	Point	k_i
$g(V) = u_1^2 - u_2 = 0$	(1,1)	$k_1 = 0.1788$

Table 1.
 Main curvature of the parabola of **Figure 5** at the point (1,1).

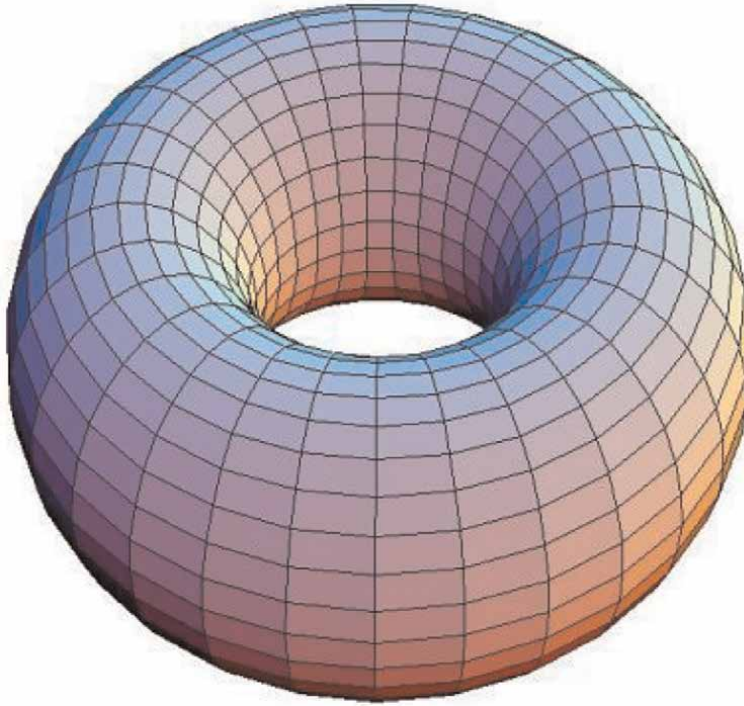


Figure 6.
Torus surface. Source: The author.

Implicit function	Point	k_i
$g(\mathbf{V}) = \sqrt{9 - (-8 + \sqrt{u^2 + v^2})^2} - z = 0$	(7.1568, 7.1568, 2.1214)	$k_1 = -0.0698$ $k_2 = -0.3333$

Table 2.
Main curvatures at four points on the torus of **Figure 6**.

$$g(\mathbf{V}) = \sqrt{9 - (-8 + \sqrt{u^2 + v^2})^2} - z = 0 \quad (13)$$

where

$$\mathbf{V} = (u, v, z) \quad (14)$$

The results of the main curvatures are shown in **Table 2** and agree with those presented in Ref. [15].

2.4.3 Main curvatures of a hyperparaboloid in standard normal space

For limit state obtained from Ref. [16], the limit state surface function of a hyperparaboloid, composed of its reliability index β and its main curvatures k_j , i.e.,

$$g(\mathbf{V}) = \beta + 0,5 \sum_{j=1}^9 k_j X_j^2 - X_{10} = 0 \quad (15)$$

where $\beta=3$, $k_1 = 0.22$, $k_2 = 0.23$, $k_3 = 0.24$, $k_4 = 0.25$, $k_5 = 0.26$, $k_6 = 0.27$, $k_7 = 0.28$, $k_8 = 0.29$, and $k_9 = 0.30$.
 and

$$\mathbf{V} = (X_1, X_2, X_3, X_4, X_5, X_6, X_7, X_8, X_9, X_{10}) \quad (16)$$

The characteristics of the (mutually independent) random variables are summarized in **Table 3**.

Determine the main curvatures of the hyperparaboloid, whose implicit function corresponds to Eq. (15), at point (0, 0, 0, 0, 0, 0, 0, 0, 0, 3).

The results obtained are presented in **Table 4**.

The results of the main curvatures are shown in **Table 4** and agree with those presented in Ref. [16].

2.4.4 Main curvatures of a hypersurface: Case of bearing capacity of a shallow footing

In the Ref. [1], reliability analysis was performed to verify the structural safety of a shallow foundation. This is a foundation supported on a homogeneous layer of silty sand subjected to the loading illustrated in **Figure 7**.

with $B = 5.0$ m, $L = 25.0$ m, $D = 1.8$ m, and $h = 2.5$ m. The limit state is ruled by the limit state surface function, whose equation is

$$G(\mathbf{U}) = q_{ult} - q = 0 \quad (17)$$

Variable	Distribution	Mean	Standard deviation
X_1, \dots, X_{10}	Standard Normal	0	1

Table 3.
 Statistical characteristics of random variables.

Implicit function	k_j
$g((\mathbf{V}) = \beta + 0,5 \sum_{j=1}^9 k_j X_j^2 - X_{10} = 0$	$k_1 = 0.22$
	$k_2 = 0.23$
	$k_3 = 0.24$
	$k_4 = 0.25$
	$k_5 = 0.26$
	$k_6 = 0.27$
	$k_7 = 0.28$
	$k_8 = 0.29$
	$k_9 = 0.30$

Table 4.
 Main curvatures of the limit state surface at point p .

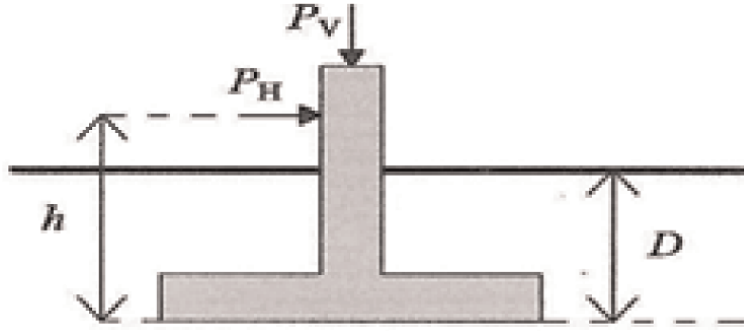


Figure 7. Shallow footing with rectangular base of width B and length L supported on a silty sand layer. Source: Ref. [10].

where q_{ult} is the vertical resistance to tipping, and q is the applied vertical pressure. The vector \mathbf{U} represents the random variables—soil cohesion (c'), soil friction angle (ϕ'), soil unit weight (γ), horizontal distributed load (P_H), and vertical distributed load (P_V), which are considered to be normally distributed.

$$q_{ult} = c' N_c s_c i_c + \gamma D N_q s_q i_q + 0.5 \gamma B' N_\gamma s_\gamma i_\gamma \quad (18)$$

where $N_q = e^{(\pi \tan \phi')} \tan^2[45^\circ + (\phi'/2)]$, $N_c = (N_q - 1) \cot \phi'$, $N_\gamma = 2 (N_q - 1) \tan \phi'$ are the (dimensionless) resistance factors, which depend on the soil friction angle, $s_q = 1 + (B' / L') \tan \phi'$, $s_c = (s_q N_q - 1) / (N_q - 1)$, and $s_\gamma = 1 - 0.3 (B' / L')$, which are also dimensionless, are factors related to the shape of the base ($B \times L$) and the eccentricity of the loads, $B' = B - 2 e_B$, $L' = L$, $e_B = h (P_H / P_V)$. Similarly, $i_q = \{1 - [P_H / (P_V + B' L' c' \cot \phi')]\}^m$, $i_c = i_q - [(1 - i_q) / N_c \tan \phi']$, and $i_\gamma = \{1 - [P_H / (P_V + B' L' c' \cot \phi')]\}^{m+1}$ are dimensionless correction factors responsible for the slope of the resulting load and $m = [2 + (B' / L')] / [1 + (B' / L')]$. Employing this approach, failure will occur when the applied vertical pressure, q , which can be calculated as

$$q = P_V / B' \quad (19)$$

becomes greater than the value of the overturning resistance, q_{ult} , calculated according to Eq. (18).

The characteristics of the random variables, which are correlated as indicated below in the matrix $\mathbf{\Omega}$, are summarized in **Table 5**.

Correlation matrix ($\mathbf{\Omega}$)

	c'	ϕ'	γ	P_H	P_V
c' (KPa)	1	-0.5	0	0	0
ϕ' (°)	-0.5	1	0.5	0	0
γ (KN/m ³)	0	0.5	1	0	0
P_H (KN/m)	0	0	0	1	0.5
P_V (KN/m)	0	0	0	0.5	1

Distribution	Variable	Mean (μ)	Standard deviation(σ)
Normal	c' (KPa)	15	4.5
Normal	φ' ($^{\circ}$)	25	5
Normal	γ (KN/m ³)	20	2
Normal	P_H (KN/m)	400	40
Normal	P_V (KN/m)	800	80

Table 5.
 Statistical characteristics of random variables.

$$\mathbf{L} = [\text{Cholesky factorization } (\mathbf{\Omega})]^T \quad (20)$$

Lower triangular matrix (\mathbf{L}) obtained through Cholesky factorization of the matrix $\mathbf{\Omega}$.

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ -0.5 & 0.866 & 0 & 0 & 0 \\ 0 & 0.577 & 0.816 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0.5 & 0.866 \end{bmatrix}$$

The vector \mathbf{V} , shown in Eq. (21), was obtained from Ref. [1], where it underwent an orthogonal transformation to make the random variables c' , φ' , γ , P_H , and P_V , which were correlated, independent.

$$\mathbf{V} = \begin{bmatrix} c' = 4.5V_1 + 15 \\ \varphi' = -0.043633231299858V_1 + 0.075574973509759V_2 + 0.436332312998582 \\ \gamma = 1.154700538379252V_2 + 1.632993161855452V_3 + 20 \\ PH = 40V_4 + 400 \\ PV = 40V_4 + 69.282032302755098V_5 + 800 \end{bmatrix} \quad (21)$$

In Ref. [1], the Cholesky factorization was also performed according to Eq. (20) in order to find the point of interest (\mathbf{V}^*) for determining the main curvatures (see **Table 6**).

Now is applied the proposed procedure to calculate the principal curvatures of the hypersurface at point $\mathbf{p} = \mathbf{V}^*$. As recommended in the proposed procedure, any random variable of the surface $g(\mathbf{V}) = 0$, that is explicitness, can be chosen. The

Variable	U^*	$V^{*} = (U^* - \mu) / \sigma$	$V^* = L^{-1} V^{**}$
c' (KPa)	14.915	-0.019	-0.019
φ' ($^{\circ}$)	18.497	-1.301	-1.514
γ (KN/m ³)	17.934	-1.033	-0.195
P_H (KN/m)	422.600	0.565	0.565
P_V (KN/m)	808.400	0.105	-0.205

Table 6.
 Obtaining the point of interest to perform the reliability analysis.

k_i
–0.0886
–0.0574
–0.0039
0.0090

Table 7.
Main curvatures of limit state surface at point \mathbf{V}^* .

surface $g(\mathbf{V}) = 0$ will be obtained by replacing the independent variables of the vector \mathbf{V} , according to Eq. (17), that is

$$g(\mathbf{V}) = c' N_c s_c i_c + \gamma D N_q s_q i_q + 0,5 \gamma B' N_\gamma s_\gamma i_\gamma - (P_V/B') = 0 \quad (22)$$

the random variables of Eq. (22) are replaced by the random variables of the vector \mathbf{V} of Eq. (21). In order to determine the function $f(\mathbf{V})$, the variable V_3 , is chosen, and then, according to what was preconceived, is made explicit. After making $V_3 = f(\mathbf{V})$ only need to determine the main curvatures of the hypersurface $f(\mathbf{V})$ at point $p = \mathbf{V}^*$ (see Table 7).

The results of the limit state surface main curvatures are accurate and in agreement with the values presented by the Ref. [10].

3. Conclusions

Five examples were solved, being three (one with an analytical solution and two via MATLAB software) from the mathematics area, aiming to validate the proposed procedure, and the last two (via MATLAB software) from the engineering area, whose principal curvatures needed to be calculated to perform the second-order reliability analysis via second-order reliability method by differential geometry (SORM DG), as recommended by Ref. [1]. However, the focus here is on generalizing the calculation of the principal curvatures at any point p , this was demonstrated through the solution of the aforementioned examples, since the procedure calculated the main curvature of a plane curve, whose implicit function has $n = 2$ (Example 2.4. 1—parabola of the second degree), the main curvatures of surfaces with implicit functions with $n = 3$ (Example 2.3—Sphere and Example 2.4.2—Torus) and the main curvatures of hypersurfaces with implicit functions with $n > 3$ (Example 2.4.3—Hyperparaboloid and Example 2.44—Shallow footing) as initially established. The procedure can be applied to any surface (or hypersurface) that is orientable and whose implicit function ($g(\mathbf{V}) = 0$) has at least one variable that can be made explicit, so that the function $f(\mathbf{V})$ can be obtained. This procedure, presented in this chapter, has been successfully applied to engineering problems requiring the calculation of the main curvatures of hypersurfaces at a point of interest, as illustrated in the simpler Example 2.43 and the more complex Example 2.44.

The mathematical procedure proposed here for computing main curvatures has proven to be very advantageous in the area of reliability analysis, including not only in terms of accuracy but also computational efficiency (e.g., see [1]).

In the area of mathematics, the procedure contributes, since it performs the calculation of the principal curvatures of plane curves and any orientable surface in space

\mathbb{R}^n , taking into account that in the literature of mathematics it is more common to calculate these curvatures up to space \mathbb{R}^3 .

Conflict of interest


The author declares no conflict of interest.

Author details

Emmanoel Ferreira
IFES—Instituto Federal do Espírito Santo, Vila Velha, Brazil

*Address all correspondence to: emmguasti@gmail.com

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Ferreira E, Freitas M, Rocha J, Sisquini G. SORM DG—An efficient SORM based on differential geometry. *REM International Engineering Journal*. 2019;**589**:60-672. DOI: 10.1590/0370-44672018720171
- [2] Lee H. *Programming and Engineering Computing with MATLAB 2021*. 1st ed. Mission: SDC Publications; 2021. p. 532
- [3] Cho S. Probabilistic stability analysis of slopes using the ANN-based response surface. *Computers and Geotechnics*. 2009;**36**:787-797. DOI: 10.1016/j.compgeo.2009.01.003
- [4] Lü Q, Low B. Probabilistic analysis of underground rock excavations using response surface method and SORM. *Computers and Geotechnics*. 2011;**2011**: 1008-1021. DOI: 10.1016/j.compgeo.2011.07.003
- [5] Lü Q, Sun H-Y, Low B. Reliability analysis of ground-support interaction in circular tunnels using the response surface method. *International Journal of Rocks Mechanics and Mining Sciences*. 2011;**48**:1329-1343. DOI: 10.1016/j.ijrmms.2011.09.020
- [6] Lü Q, Chan L, Low B. Probabilistic evaluation of ground-support interaction for deep rock excavation using artificial neural network and uniform design. *Tunneling and Underground Space Technology*. 2012;**32**:1-18. DOI: 10.1016/j.tust.2012.04.014
- [7] Chan L, Low B. Practical second-order reliability analysis applied to foundation engineering. *International Journal for Numerical and Analytical Methods in Geomechanics*. 2012;**36**:1387-1409. DOI: 10.1002/NAG.1057
- [8] Zeng P, Jimenez R. An approximation to the reliability of series geotechnical systems using a linearization approach. *Computers and Geotechnics*. 2014;**62**: 304-309. DOI: 10.1016/J.COMPGeo.2014.08.007
- [9] Zeng P, Jimenez R, Jurado-Piña R. System reliability analysis of layered soil slopes using fully specified slip surfaces and genetic algorithms. *Engineering Geology*. 2015;**193**:106-117. DOI: 10.1016/J.ENGGeo.2015.04.026
- [10] Zeng P, Jimenez R, Li T. An efficient quase-Newton approximation-based SORM to estimate the reliability analysis of geotechnical problems. *Computers and Geotechnics*. 2016;**76**:33-42. DOI: 10.1016/J.COMPGeo.2016.02.003
- [11] Zeng P, Li T, Jimenez R, Freng X, Chen Y. Extension of quase-Newton approximation-based SORM for series system reliability analysis of geotechnical problems. *Engineering with Computers*. 2017;**34**:215-224. DOI: 10.1007/s00366-017-0536-8
- [12] Rodrigues P. *Introdução às curvas e superfícies*. 1st ed. Niterói: EdUFF; 2001. p. 236
- [13] Carmo M. *Differential Geometry of Curves and Surfaces: Revised and Updated*. 2nd ed. Mineola: Dover Publications; 2016. p. 528
- [14] Araújo P. *Geometria diferencial*. 1st ed. Rio de Janeiro: IMPA; 1998
- [15] Gray A. *Modern Differential Geometry*. 3rd ed. Boca Raton: CRC Press; 2006. p. 1053
- [16] Kiureghian A, De Stefano M. *An Efficient Algorithm for Second-Order Reliability Analysis*. Report n°. UCB/SEMM – 90/20. University of California; 1990

Section 3

Modern Research Topics
in Topology

Topology Optimization Generating System

Sandra Nadim Aldarrouj and Ibrahim Alshannour

Abstract

In this project, we started with discussing the structural topology approaches in general throughout the concept of topology optimization. We discussed the different types of topologies and their mathematical functions, and then, we established the idea of creating the topology optimization generating system. We made a collaboration of multiple programs and algorithms to create a friendly user environment for mechanical engineers mostly to design their own work with the best optimized solution possible with the less available volume of material. Afterwards they can test the result with a physics simulator and apply any constraints and physical force anywhere they want.

Keywords: meta-heuristics methods, density based, stiffness matrices, finite state elements method, topology optimization

1. Introduction

The structural optimization is considered to be a mechanical designing to be a mechanical designing concept. The main purpose behind it is to find the optimal type of structure considering the constraints applied on the structure target.

2. Optimization problem

The main concept behind the topology optimization problems is to find the variable x where it minimizes or maximizes the objective function $f(x)$ considering the constraints applied to the initial structure. There are many methods to solve this problem by using an optimization model to solve our problem.

Every optimization model has the following [1]:

- **Objective function $f(x)$:**

This function describes the structure goal that the user wants to optimize by finding the region of feasible solutions. Thus, the objective function job is to minimize or maximize the amount of manufacturing material in every micro-structure.

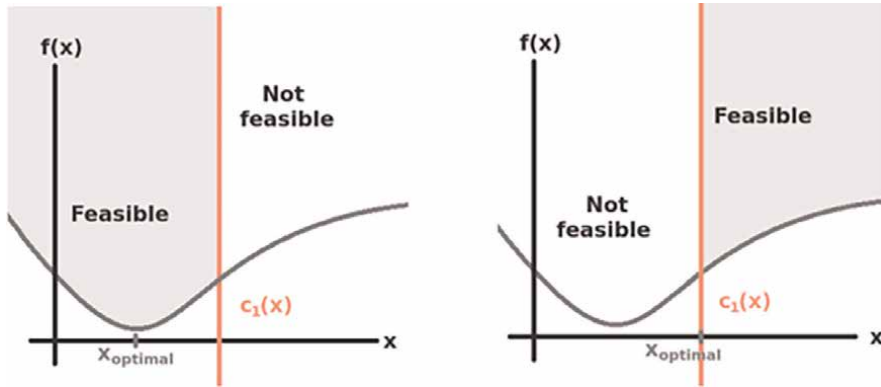


Figure 1.
Feasibility regions.

- **Design variable (x):**

The variable (x) describes the parameters that the objective function depends on them, whereas the objective function searches for the best results to reach the optimal solution for the structure goal.

- **Constrains:**

Having constrains in the optimization problems is conditional where it depends on each individual issue, but in real-life optimization problems, there must be few constrains to mark the feasible regions.

Figure 1 shows how the feasible regions can be different due to each optimization problem.

3. Structural optimization

The main idea behind the structural optimization is making an assemblage of materials to sustain loads in the best way possible, and thus, it is considered to be a mechanical designing process that aims to find the optimized structure considering the designing area [2].

4. The main types of structural optimization

- **Size Optimization:**

Size optimization is considered to be the simplest method where the optimization parameters are given by the structural parameters like height, width thickness and corners. In this method, the shape of the target is known, but the goal is to find a better shape. In this type, the structural elements are the design variables, and the objective function aims to find different shapes of the structure goal [2].

- **Shape Optimization:**

This type is considered to be an extension of size optimization, whereas it allows us to set the range of the elements of material interdependence, and the goal of

this method is to change rearrange the place of elements without changing the interdependency of the elements themselves. In this type, the parameters of the objective function are the coordinates of the structure corners considering few parameters to build the edges of the structure target like reducing the stress—decrease or increase the exertion applied on the structure ... etc.

- Topology Optimization:

This type is considered to be better than the other types because it does not apply constraints on the structure goal. In this type the design target can be partitioned into elements whereas the optimization parameters are a group of elements, these elements are the material parameters that shows the material variation where the structure goal has holes in it and where it does not. The objective function is to find the best way to vary the material.

5. Finite element method

This method is used to find a numeric solution for the differential equations. The main purpose behind solving them is to describe the behavior of any physical phenomenon, whereas we try to solve the elastic problem described by the displacement as a result of applying any force on the elastic initialized structure. The elasticity of the initialized structure can be static at certain elements. The main role of the finite element method is being linked with the topology optimization objective function and its constraints [3, 4].

In this method, we turn the initial structure into small pieces called the micro-structures, whereas any physical force applied on the initialized structure will have an effect on the micro-structures.

- Linear elastic problem:

We implement the displacement field in the stress and strain problems as a two-way displacement field for each element in the designing area.

The displacement vector is represented as

$$U = U(X, Y) = \begin{bmatrix} u(x, y) \\ v(x, y) \end{bmatrix} = \begin{bmatrix} u \\ v \end{bmatrix} \quad (1)$$

Whereas each u and v represent the displacement on x , y coordinates at a specific element.

The strain can be calculated using the displacement vectors on the plane xy using the general elastic theory, whereas we implement the strain vector as.

$$\varepsilon = \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \gamma_{xy} \end{bmatrix} = \begin{bmatrix} \frac{\partial u}{\partial x} \\ \frac{\partial v}{\partial y} \\ \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \end{bmatrix} = \begin{bmatrix} \frac{\partial}{\partial x} & 0 \\ 0 & \frac{\partial}{\partial y} \\ \frac{\partial}{\partial y} & \frac{\partial}{\partial x} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = Bu \quad (2)$$

Whereas each ε_x and ε_y represent the normal strain on x, y coordinates, and γ_{xy} represents the shear strain on the plane xy.

The stress vector:

$$\sigma = \begin{bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{bmatrix} \quad (3)$$

Whereas each σ_x and σ_y represent the normal stress on x, y coordinates, and γ_{xy} represents the tangential stress on the plane xy.

The relationship between stress and strain represented as

$$\sigma = D \varepsilon \quad (4)$$

In the general status:

$$\sigma = D(\varepsilon - \varepsilon_0) + \sigma_0 \quad (5)$$

This equation reserves on the linear behavior until a certain limit called the elastic limit.

In the previous equation, D represents the basic matrix that has all the elasticity properties of the manufacturing material:

$$D = \begin{bmatrix} d_{11} & d_{12} & 0 \\ d_{21} & d_{22} & 0 \\ 0 & 0 & d_{33} \end{bmatrix} \quad (6)$$

Based on Maxwell–Betti theory, then D is symmetrical [5].

Using isotropic material, D can be calculated for stress as

$$d_{11} = d_{22} = \frac{E}{1 - \nu^2} \quad (7)$$

$$d_{12} = d_{21} = \nu d_{11} \quad (8)$$

$$d_{33} = \frac{E}{2(1 + \nu)} \quad (9)$$

The strain is calculated by

$$d_{11} = d_{22} = \frac{E(1 - \nu)}{(1 + \nu)(1 - 2\nu)} \quad (10)$$

$$d_{12} = d_{21} = d_{11} \frac{\nu}{1 - \nu} \quad (11)$$

$$d_{33} = \frac{E}{2(1 + \nu)} \quad (12)$$

Whereas E represents Young's modulus, and ν represents the Poisson radius.

- Discrete Solution:

This equation allows to represent the effect of the generated work done by the internal and external forces of the design and links the stress, strain and pressure:

$$\begin{aligned} \iint_A (\delta \epsilon_x \sigma_x + \delta \epsilon_y \sigma_y + \delta \gamma_{xy} \tau_{xy}) h dA = \iint_A (\delta u b_x + \delta v b_y) h dA \\ + \iint_s (\delta u t_x + \delta v t_y) h ds + \int_i (\delta u_i U_i + \delta v_i V_i) \end{aligned} \quad (13)$$

This can be written in a matrix way:

$$\iint_A \delta \epsilon^T \sigma h dA = \iint_A \delta u^T b h dA + \int_s \delta u^T t h ds + \sum_i \delta u_i^T q_i \quad (14)$$

Whereas $\delta u, \delta \epsilon, \sigma$ represent the theoretical vectors, the theoretical stress and the pressure fields.

h represents the thickness of the initial structure.

b, t, q represent the forces applied on the initial structure given by:

The forces applied on the structure elements:

$$b = \begin{bmatrix} b_x \\ b_y \end{bmatrix} \quad (15)$$

The forces applied on the edges:

$$t = \begin{bmatrix} t_x \\ t_y \end{bmatrix} \quad (16)$$

The forces applied on the points:

$$q = \begin{bmatrix} U \\ V_y \end{bmatrix} \quad (17)$$

All these forces will be given as constraints for the optimization process.

Using finite element method, we can solve our problem using the discrete method where we calculate the solution at some points, then we interpolate the solution in the middle points, and thus, the elements will be generated from the points and the links between them and that is how we create a complete generation for the initial structure elements as shown in **Figure 2**:

To apply the theoretical work of the elements depending on Eq. (2) that can be written as:

$$\iint_{A^e} \delta u^T \sigma h dA^e = \iint_{A^e} \delta u^T b h dA^e + \int_{s^e} \delta u^T t h ds^e + \delta u^T q^e \quad (18)$$

Now, we can partition and except the theoretical work vectors since it is randomly generated from the work of elements equation:

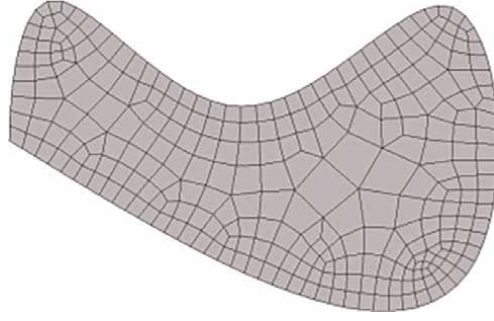


Figure 2.
Initial structure elements.

$$\iint_{A^e} B^T \sigma h dA^e = \iint_{A^e} b h dA^e + \int_{s^e} t h ds^e + q_e \quad (19)$$

Replacing the general Eq. (5) with the first part of the previous equation:

$$\iint_{A^e} B^T (D(\varepsilon - \varepsilon_0) + \sigma_0) h dA^e = \iint_{A^e} b h dA^e + \int_{s^e} t h ds^e + q_e \quad (20)$$

Algebraically, we get

$$h \iint_{A^e} B^T \varepsilon dA^e = h \iint_{A^e} B^T D \varepsilon_0 dA^e - h \iint_{A^e} B^T \sigma_0 dA^e + h \iint_{A^e} b dA^e + h \int_{s^e} t ds^e + q^e \quad (21)$$

Replacing Eq. (2) with the first part of the previous equation, we get

$$h \iint_{A^e} B^T D B dA^e u = h \iint_{A^e} B^T D \varepsilon_0 dA^e - h \iint_{A^e} B^T \sigma_0 dA^e + h \iint_{A^e} b dA^e + h \int_{s^e} t ds^e + q^e \quad (22)$$

This equation can be written as

$$K^e u^e = [f_{\varepsilon}^e - f_{\sigma}^e + f_b^e + f_s^e] + q^e = f^e + q^e \quad (23)$$

After assembling all the element wide matrices, we get this linear equation:

$$K \cdot u = f \quad (24)$$

6. Topology optimization mathematical function

We start by solving the linear equation:

$$K \cdot u = f \quad (25)$$

Whereas:

K is the stiffness matrix (determines the hardness of the material).

U is the displacement vector (determines where not to add material during design.)

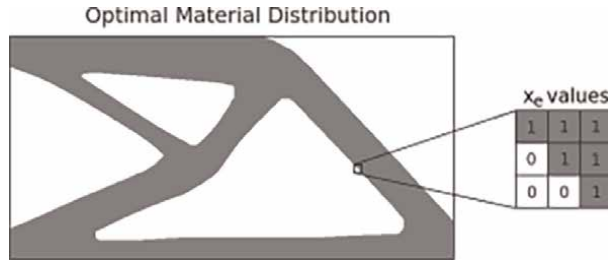


Figure 3.
Optimal Material Distribution.

f is the forces vectors.

The stiffness matrix K can be factorized as

$$K = E.\hat{K} \quad (26)$$

Whereas:

E is the Young's parameter (determines the solidity of the structure).

\hat{K} is the factorized stiffness matrix.

The main purpose behind dividing the stiffness matrix is to divide the structure goal into small parts to determine which element is more important than others. Now, we can factorize the elements using the equation:

$$k_e = E.\hat{k}_e \quad (27)$$

Whereas:

E is the Young's parameter for each element e .

Now, we can link each element with any of the parameters of designing with each parameter from the constrains using this equation:

$$E_e = x_e. E \quad (28)$$

Replacing E_e , we get:

$$k_e = x_e. E.\hat{k}_e \quad (29)$$

Thus, each variable x_e is considered to be a parameter for the topology optimization problems. When $x_e = 1$, we get material, when $x_e = 0$, we get an empty hole. Some methods use binary design variables, whereas others may use continuous values between 0 and 1.

Figure 3 shows how assigning the optimal values to the variables in topology optimization methods can give the best results:

7. Topology optimization approaches

- Density-Based Method (Gradient Methods) [6]:

This method depends on creating a discrete designing space using solid elements or solid nodes, whereas it is available to change the volume of design or the

stiffness of each micro-structure by changing the value of the variable x_e considering its range $[x_{\min}, x_{\max}]$, and thus, we can create a physical simulation by changing the density or the solidarity (Young's parameter).

- Solid Isotropic Material with Penalization (SIMP) [7]:

This methodology is also gradient-based where it preserves on the solidity of the material. The structure goal should get fragmented into micro-structures so that the user can delete any unwanted parts from the initial structure, and in this method, the stiffness stays the same for each element, whereas we make several alignments to get to the structure goal. The algorithm of this method starts with setting the stiffness of all elements elected to be the same taking in consideration that all the selected elements have the same volume, and thus, the algorithm will frequently start setting the optimality of each element considering the constrains needed to be applied on the structure goal.

This method was used in many researches like [5, 8, 9].

SIMP algorithm steps:

1. Create initial structure fragments.
2. Set the micro-structure to be solid elements gradually.
3. Analyze the elements considering the physical constrains applied (forces, stress ... etc).
4. Find the optimal value for each element considering how important is the element for the optimization process.
5. Reduce the number of elements frequently.
6. Use the “soft kill” penalization concept where each element has its own situation of material (has material—space—partly has material).

The pros of SIMP algorithm:

1. The initial structure does not have to be homogenous.
2. Considered to be mathematically functional.
3. Considered to be approximately flexible with any design.
4. Does not require hard mathematics understanding.
5. Can be used with any physics constrains.

The cons of SIMP algorithm:

1. Depends on the mesh of the material.

- 2. There might be elements with two types of materials.
- 3. Depends on penalization.

Figure 4 shows how SIMP algorithm works:

- Rational Approximation of Material Properties [7, 10]:

This method has been established to fix the designs that takes the weather circumstances in consideration like: wind force, humidity, snow ... etc., and the main concept behind it is to be an initial alternative formula of (SIMP) algorithm where the designer determines the designing void area as a hydro-static liquid unable to be pressured, the liquid transforms the pressure force without leaving any marks on the surface.

This method was used in many researches like [11–13].

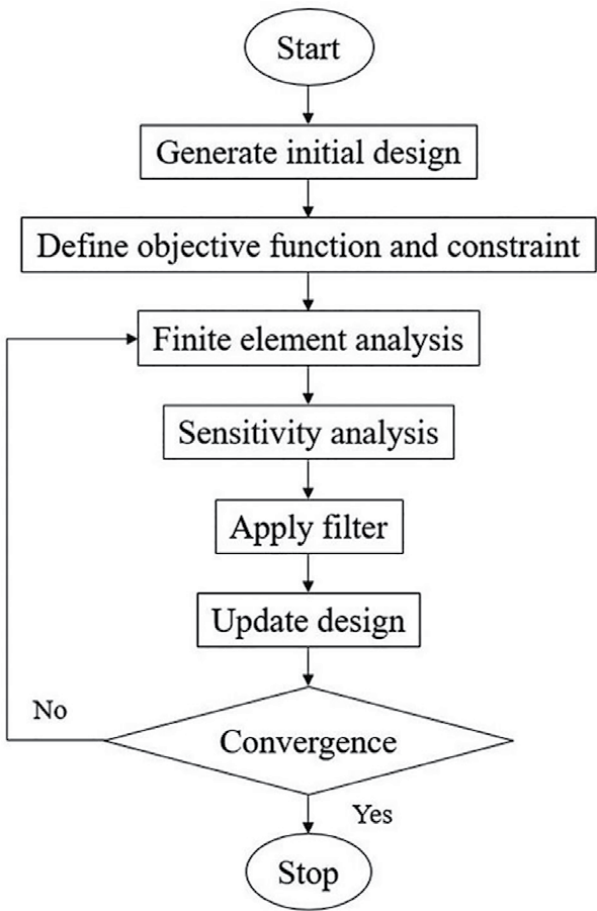


Figure 4.
SIMP Algorithm steps.

The cons of RAMP:

- Depends on penalization.
- It shows hard mathematical problems when using low-stiffness structured elements.

Figure 5 clarifies the idea of RIMP:

Heuristic Methods (Discrete Methods) [6]:

This type depends on the knowledge of the user to get to the structure goal, and it is not necessary to have mathematical knowledge base. These methods considered to be simple, but on the other hand, reaching the local or global optimal results is not guaranteed because it does not change any of the variables or parameters of the structure goal considering the optimization algorithm applied on the initial structure.

These methods were used in many researches like [14–16].

Meta-Heuristic Methods [6]:

These methods considered to be randomized, and in many topology optimizing cases, these methods depend on population-based proof of concept where they are called evolutionary algorithms. Each element of structure is being evaluated by a randomizing solution, whereas the best elements are picked up by the algorithm to create the best evaluating solution. The main steps of these algorithms are selection, crossover, mutation and replacement applied on the chosen elements of structure.

These methods were used in many researches like (**Figure 6**) [17–19] .

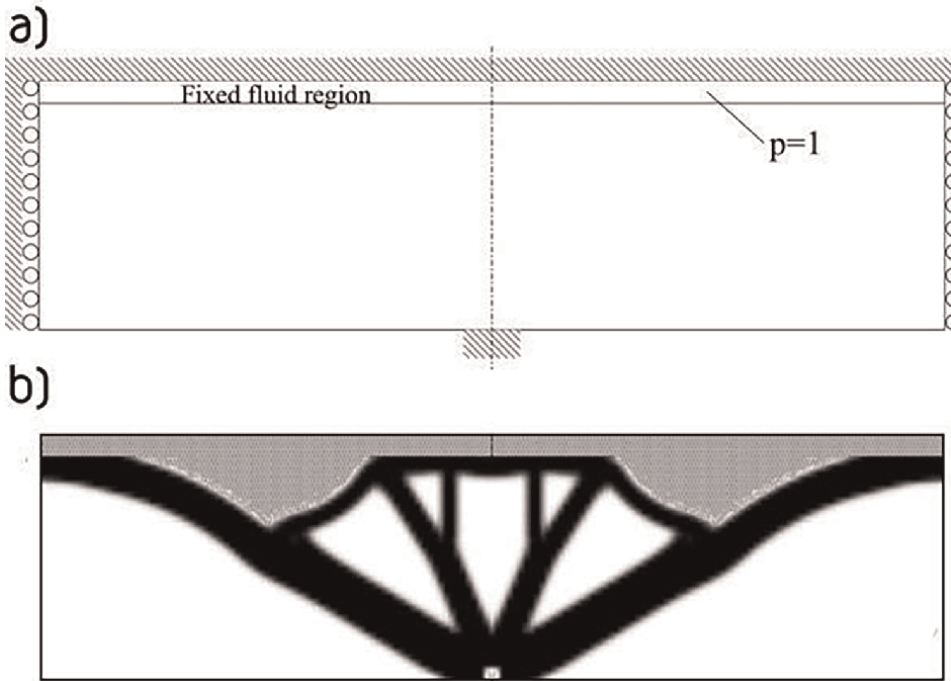


Figure 5.
RIMP Algorithm result.

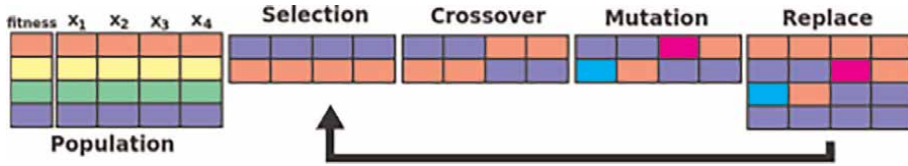


Figure 6.
 Meta Heuristic methods work.

The pros of meta-heuristic methods:

- Capable of avoiding local optimal limitations.
- Capable of being approached from global optimal limitation.
- Does not need a lot of information about the initial structure.

The cons of meta-heuristic methods:

- Cannot handle a lot of variables about the initial structure.
- Require heavy types of computing resources.

8. Topology optimization generating system

The main purpose that stands behind this project is the need of reducing the amount of material used in manufacturing without weakening the structure goal, also making the designing process easier by adding the constrains strictions and physical behavior so that the structure goal became more mesh-linked and efficient. This project stands upon finding the best variation of the material of manufacturing in strict circumstances so that the structure goal has the same efficiency of the initial structure but less manufacturing material.

9. The functional requirements

- Choosing the initial structure:

Using FreeCAD software, we managed to create a designer-friendly environment that gives the users all the freedom during their journey of designing the initial structure, whereas they can design their 3D model in any shape, volume and all the details they want or import a previous model they already have and apply their changes on it.

- Initial goal partition:

In this part, we change the initial structure into a solid structure, and then, we applied the partitioning method that is called the finite elements method where the user can change then approve the method's work.

- Apply the physical constraints:

Using a plugin tool called “Calculix”, the designer can apply any type of physical constraints that he/she wants, like: material type, stiffness ... etc.

- Apply the functional constraints:

While using the plugin “Calculix”, the designer can apply the fixed points and the force vectors parameters like: the transformation and rotation of the vector and the intensity of force ... etc.

- Managing the evacuation process:

After applying all the constraints needed, the 3D model file would be an input file to the evacuation process with another parameters that the algorithm needs like the iteration number whether or not save all iterations ... etc.

For this level, we managed to implement the SIMP algorithm as a plugin to FreeCAD to create a user-friendly interface to manage all the parameters and constraints needed to be applied on the initial 3D model.

The user interface includes:

- a. Choose the path of saving all the iterated models that have been evacuated.
- b. Set all the parameters of the SIMP algorithm.
- c. Start the evacuation process.
- d. Test the results on the physical simulator that will be explained later on.
- e. Show the results of the most optimized 3D model.
- f. Represent a GIF of all the iterations of the evacuation process.
- g. Test the results using a physics simulator.

- The evacuation process works:

Using the solid isotropic material with penalization (SIMP) algorithm to merge and solve the problem, whereas it is a homogeneous method that allows to give each element of the partitioned initial structure its own value x_e given by the range $[x_{min}, x_{max}]$ whereas $x_{min} = 1$ (and)

$$x_{max} \approx 1E - 3.$$

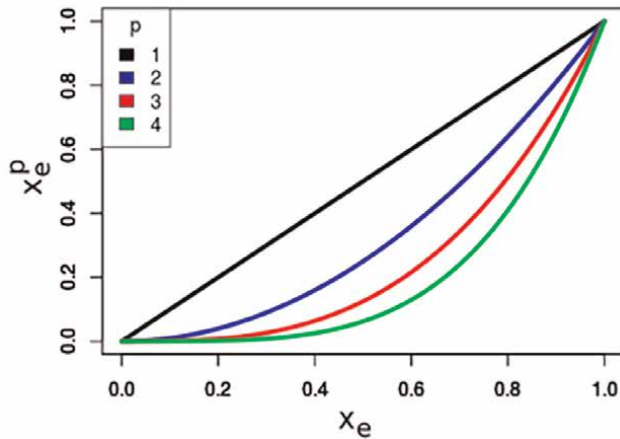


Figure 7.
Evacuation Process work.

This value represents the amount of manufacturing material inside these element's structure so that it allows to simulate the solidity of the structure by varying the material during the evacuation process.

The stiffness of the material and Young's variable can be changed using the equations:

$$E_e = x_e^p \cdot E \quad (30)$$

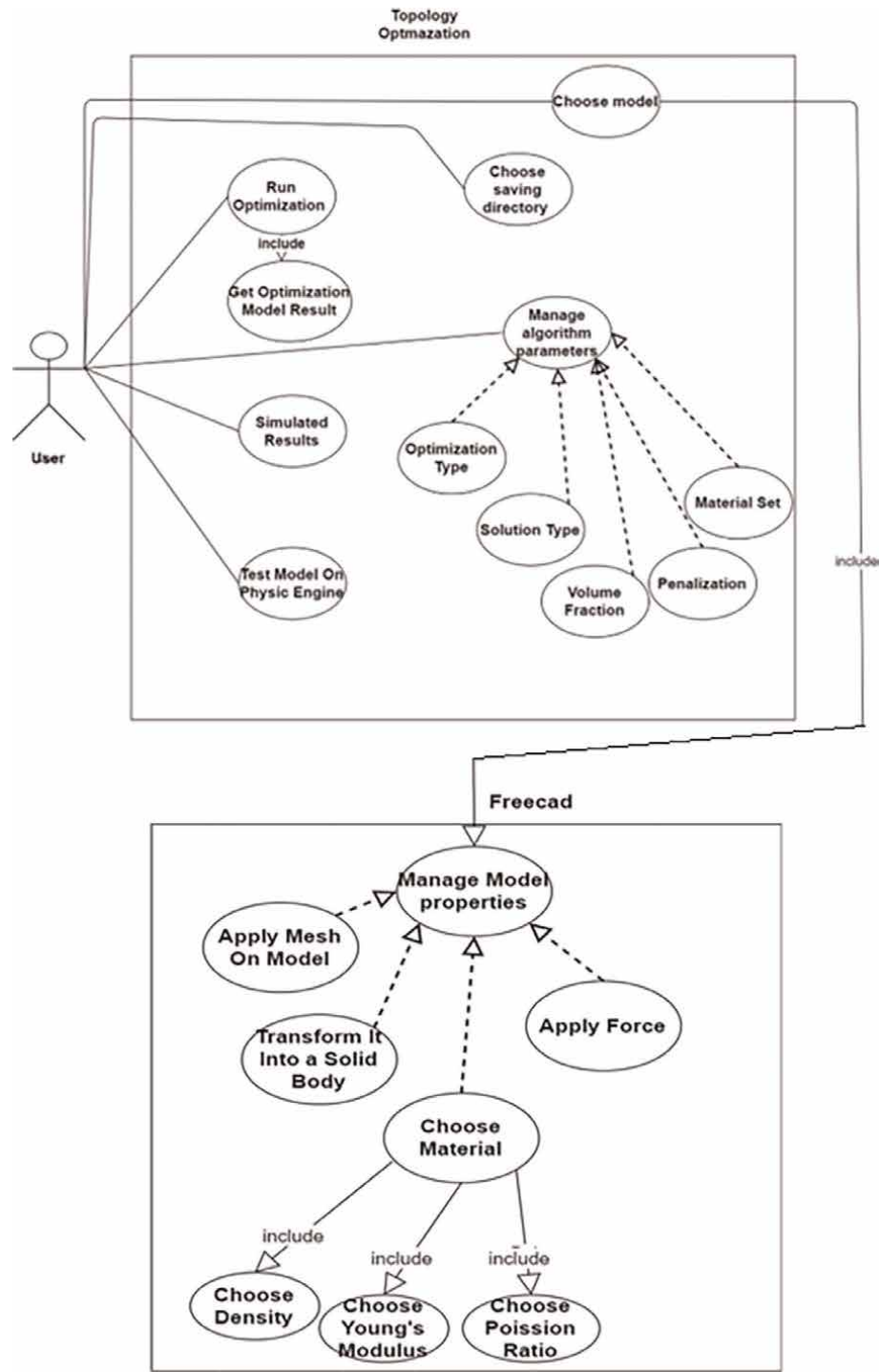
$$\rho_e = x_e^p \cdot \rho \quad (31)$$

The parameter p is called the power penalization, whereas it aims to penalize the average values of x_e . Thus, it leads to the polynomial penalization as shown in the **Figure 7**, whereas the result would be small changes in the variable x_e of the material solidity and stiffness proportion inside each element of the micro-structure.

- The physical simulator:

We managed to build the physical simulator using the game engine Unity 2020.3.5f1, whereas it has the physics engine called Nvidia Physx 3.3 built by the famous Nvidia company. The physics simulator has a user interface that allows the user to choose the evacuated 3D model from the previous level in any format: fbx, obj, stl, and dae and then adds the applied constraints from the earlier stage, and afterwards the user can test the model by adding linear or torque force to see how the mesh would be deformed.

10. Use case diagram



11. Conclusion


The methodology that has been used is considered to be simple and effective for the topology optimization problems especially the problems that need to reduce optimization over the volume constraints where it evaluates the structure frequently and demands hundreds of frequencies and that is less 5% than the evolutionary algorithms. On the other hand, restricting the SIMP algorithm with a certain number of frequencies might not give it the freedom to reach the optimal solution so that the user should be careful how to implement the variables of the algorithm.

Author details

Sandra Nadim Aldarrouj* and Ibrahim Alshannour
Damascus University, Damascus, Syria

*Address all correspondence to: sandradarrouj@gmail.com

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Nocedal J, Wright SJ. Numerical Optimization. 2nd ed. Springer; 2006. Available from: <https://link.springer.com/book/10.1007/978-0-387-40065-5>
- [2] Pena SIV, Rionda SB. Topology Optimization Algorithms for the Solution of Compliance and Volume Problems in 2D. Mexico: A.C. Guanajuato; 2016. p. 24
- [3] Botello S, Esqueda H, Gomez F, Moreles M, Onate E. Modulo de aplicaciones del metodo de los elementos finitos mefi 1.0, chapter manual teorico. CIMAT & CIMNE; 2004. pp. 6-69. Available from: <https://cimat.repositorioinstitucional.mx/jspui/bitstream/1008/523/1/TE%20605.pdf>
- [4] OC Zienkiewicz, RL Taylor, and JZ Zhu. The Finite Element Method: Its Basis and Fundamentals. 2005. Available from: https://www.researchgate.net/publication/334974685_State_of_the_art_of_generative_design_and_topology_optimization_and_potential_research_needs
- [5] Bendsoe MP. Optimal shape design as a material distribution problem. Structural Optimization. 1989;1(4):193-202
- [6] Pena SIV, Rionda SB. Topology Optimization Algorithms for the Solution of Compliance and Volume Problems in 2D. Mexico: A.C. Guanajuato; 2016. pp. 4-6
- [7] Tyflopoulos E, Flem DT, Steinert M, Olsen A. State of the Art of Generative Design and Topology Optimization and Potential Research Needs. Linköping, Sweden; 2018. pp. 6-8
- [8] Rozvany GIN. Aims, scope, methods, history and unified terminology of computeraided topology optimization in structural mechanics. Structural and Multidisciplinary Optimization. 2001; 21(2):90-108
- [9] Zhou M, Rozvany GIN. The Coc Algorithm .2. Topological, Geometrical and Generalized Shape Optimization. Computer Methods in Applied Mechanics and Engineering. 1991;89 (1-3):309-336
- [10] Nha Chu D, Xie YM, Hira A, Steven GP. On various aspects of evolutionary structural optimization for problems with stiffness constraints. Finite Elements in Analysis and Design. 1997;24(4):197-212
- [11] Stolpe M, Svanberg K. An alternative interpolation scheme for minimum compliance topology optimization. Structural and Multidisciplinary Optimization. 2001;22(2):116-124
- [12] Deaton JD, Grandhi RV. A survey of structural and multidisciplinary continuum topology optimization: Post 2000. Structural and Multidisciplinary Optimization. 2014; 49(1):1-38
- [13] Luo Z, Chen L, Yang J, Zhang Y, Abdel-Malek K. Compliant mechanism design using multi-objective topology optimization scheme of continuum structures. Structural and Multidisciplinary Optimization. 2005; 30(2):142-154
- [14] Zhou M, Rozvany G. On the validity of ESO type methods in topology optimization. Structural and Multidisciplinary Optimization. 2001; 21(1):80-83
- [15] Xie YM, Steven GP. A simple evolutionary procedure for structural

optimization. *Computers & Structures*. 1993;**49**(5):885-896

[16] Xie YM, Huang X. Recent developments in evolutionary structural optimization (ESO) for continuum structures. In: *IOP Conference Series: Materials Science and Engineering*. 2010. Available from: https://www.researchgate.net/publication/230941501_Recent_developments_in_evolutionary_structural_optimization_ESO_for_continuum_structures

[17] Balamurugan R, Ramakrishnan CV, Singh N. Performance evaluation of a two stage adaptive genetic algorithm (tsaga) in structural topology optimization. *Applied Soft Computing*. 2008;**8**(4):1607-1624

[18] Bureerat S, Limtragool J. Performance enhancement of evolutionary search for structural topology optimisation. *Finite Elements in Analysis and Design*. 2006;**42**(6): 547-566

[19] Hansel W, Treptow A, Becker W, Freisleben B. A heuristic and a genetic topology optimization algorithm for weight- minimal laminate structures. *Composite Structures*. 2002;**58**(2): 287-294

Chapter 6

Some Recent Advances in Non-Hausdorff Topology

Xiaoquan Xu

Abstract

In the past few years, the research in non-Hausdorff topology, especially on sober spaces and well-filtered spaces, has got some breakthrough progress. In this paper, we shall present a brief summarising survey on some of such development. Some related problems are listed.

Keywords: Sober space, well-filtered space, d-space, reflection, Scott topology, Smyth power space

1. Introduction

In most topology books, the Hausdorff separation property is assumed from the very start and very little information is contained on non-Hausdorff spaces. In classical mathematics, most topological spaces are indeed Hausdorff. But non-Hausdorff spaces are important already in algebraic geometry, and crucial in fields such as domain theory. Indeed, in connection with order, non-Hausdorff spaces (especially T_0 spaces) play a more significant role than Hausdorff spaces.

Sobriety is probably the most important and useful property of T_0 spaces. With the development of domain theory, another two properties also emerged as very useful and important properties for non-Hausdorff topology¹ theory: d -space and well-filtered space (see [1–52]). In the past few years, some remarkable progresses have been achieved in understanding such structures. In this chapter, we shall make a brief survey on some of this progress and list a few related problems.

2. Preliminary

We now recall some basic concepts and notations that will be used in the paper. For further details, we refer the reader to [4, 7, 53].

A partial order \leq on a set X is a *transitive* (i.e., $x \leq y$ and $y \leq z$ imply $x \leq z$), *reflexive* (it means $x \leq x$ for any $x \in P$), and *antisymmetric* (i.e., $x \leq y$ and $y \leq x$ imply $x = y$)

¹ This research was supported by the National Natural Science Foundation of China (Nos. 12071199, 11661057).

relation. A partially ordered set, or *poset* for short, is a nonempty set P equipped with a partial order \leq .

Let P be a poset. For any $A \subseteq P$, let $\downarrow A = \{x \in P : x \leq a \text{ for some } a \in A\}$ and $\uparrow A = \{x \in P : x \geq a \text{ for some } a \in A\}$. For each $x \in P$, we write $\downarrow x$ for $\downarrow \{x\}$ and $\uparrow x$ for $\uparrow \{x\}$. A subset A of P is called a *lower set* (resp., an *upper set*) if $A = \downarrow A$ (resp., $A = \uparrow A$). Define $P^{(<\omega)} = \{F \subseteq P : F \text{ is a nonempty finite set}\}$ and **Fin** $P = \{\uparrow F : F \in P^{(<\omega)}\}$. For a nonempty subset A of P , define $\max(A) = \{a \in A : a \text{ is a maximal element of } A\}$ and $\min(A) = \{a \in A : a \text{ is a minimal element of } A\}$. The symbol \mathbb{N} will denote the poset of all natural numbers in the usual order.

A nonempty subset D of a poset P is called *directed* if every two elements in D have an upper bound in D . Let $\mathcal{D}(P)$ denote the set of all directed sets of P . A poset P is called a *directed complete poset*, or *dcpo* for short, if the supremum $\vee D$ of D exists in P for every $D \in \mathcal{D}(P)$. A subset $I \subseteq P$ is called an *ideal* if I is a directed lower subset of P . Let $\text{Id}(P)$ denote the poset of all ideals of P with the set inclusion order. Dually, we define the *filters* and denote the poset of all filters of P by $\text{Filt}(P)$.

For a poset Q , the *upper topology* on Q , generated by $\{Q \setminus \downarrow x : x \in Q\}$ as a subbase, is denoted by $\nu(Q)$. Dually, we define the *lower topology* on Q and is denoted by $\omega(Q)$. The upper sets of Q form the (*upper*) *Alexandroff topology* $\alpha(Q)$.

Definition 2.1. For a poset P and $U \subseteq P$, U is said to be *Scott open* if

- i. $U = \uparrow U$, and
- ii. for any directed subset D for which $\vee D$ exists, $\vee D \in U$ implies $D \cap U \neq \emptyset$.

All Scott open subsets of P form a topology, called the *Scott topology* on P and denoted by $\sigma(P)$. The space $\Sigma P = (P, \sigma(P))$ is called the *Scott space* of P . The topology generated by $\omega(P) \cup \sigma(P)$ is called the *Lawson topology* on P and is denoted by $\lambda(P)$. The space $\Lambda P = (P, \lambda(P))$ is called the *Lawson space* of P .

The following result is well-known (cf. [4], Proposition II-2.1).

Lemma 2.2. Let P, Q be posets and $f : P \rightarrow Q$. Then the following two conditions are equivalent:

- 1. f is *Scott continuous*, that is, $f : \Sigma P \rightarrow \Sigma Q$ is continuous.
- 2. For any $D \in \mathcal{D}(P)$ for which $\vee D$ exists, $f(\vee D) = \vee f(D)$.

For a T_0 space X , the *specialization order* \leq_X on X is defined by $x \leq_X y$ iff $x \in \overline{\{y\}}$. In the following, when a T_0 space is considered as a poset, the order shall mean the specialization order provided a different one is specified. Let $\mathcal{O}(X)$ (resp., $\mathcal{C}(X)$) be the set of all open subsets (resp., closed subsets) of space X . For $A \subseteq X$, the closure of A in X is denoted by $\text{cl } A$, or simply by \overline{A} . Define $\mathcal{S}_c(X) = \{\overline{\{x\}} : x \in X\}$ and $\mathcal{D}_c(X) = \{\overline{D} : D \in \mathcal{D}(X)\}$. For two spaces X and Y , we use the symbol $X \cong Y$ to denote that X and Y are homeomorphic.

A T_0 space X is called a *d-space* (or *monotone convergence space*) if X (with the specialization order) is a dcpo and $\mathcal{O}(X) \subseteq \sigma(X)$ (cf. [4, 32]). For any dcpo P , ΣP is clearly a *d-space*. The category of all *d-spaces* and continuous mappings is denoted by **Top_d**.

For a d -space X and a nonempty closed subset of X , if $x \in A$, then by Zorn's Lemma there is a maximal chain C_x in A with $x \in C_x$. Since X is a d -space, $c_x = \vee C_x$ exists and $c_x \in A$. By the maximality of C_x , we have $c_x \in \max(A)$ and $x \leq c_x$. Therefore, $A \subseteq \downarrow \max(A) \subseteq \downarrow A = A$, and hence $A = \downarrow \max(A)$. So we have the following.

Lemma 2.3. *Let X be a d -space. If A is a nonempty closed subset of X , then $A = \downarrow \max(A)$ and hence $\max(A) \neq \emptyset$.*

Lemma 2.4. ([39]) *Let P be a poset and Y a d -space. Then for any $f : P \rightarrow Y$, the following two conditions are equivalent:*

1. $f : \Sigma P \rightarrow Y$ is continuous.
2. $f : \Sigma P \rightarrow \Sigma Y$ is continuous.

A nonempty subset A of a T_0 space X is called *irreducible* if for any $\{F_1, F_2\} \subseteq \mathcal{C}(X)$, $A \subseteq F_1 \cup F_2$ implies $A \subseteq F_1$ or $A \subseteq F_2$. We denote by $\text{Irr}(X)$ (resp., $\text{Irr}_c(X)$) the set of all irreducible (resp., irreducible closed) subsets of X . Clearly, every subset of X that is directed under \leq_X is irreducible. The space X is called *sober*, if for any $F \in \text{Irr}_c(Y)$, there is a (unique) point $a \in X$ such that $F = \overline{\{a\}}$. The category of all sober spaces and continuous mappings is denoted by **Sob**.

For a family X_i ($i \in I$) of T_0 spaces and the product space $X = \prod_{i \in I} X_i$, let $p_i : X \rightarrow X_i$ ($i \in I$) be the i th projection.

Lemma 2.5. ([42]) *Let $X = \prod_{i \in I} X_i$ be the product space of T_0 spaces X_i ($i \in I$). If $A \in \text{Irr}_c(X)$, then $A = \prod_{i \in I} p_i(A)$ and $p_i(A) \in \text{Irr}_c(X_i)$ for each $i \in I$.*

A subset A of a T_0 space X is called *saturated* if A equals the intersection of all open sets containing it (equivalently, A is an upper set with respect to the specialization order). We use $\mathcal{K}(X)$ to denote the set of all nonempty compact saturated subsets of a T_0 space X and endow $\mathcal{K}(X)$ with the *Smyth order*, that is, for $K_1, K_2 \in \mathcal{K}(X)$, $K_1 \sqsubseteq K_2$ iff $K_2 \subseteq K_1$. Let $\text{OFilt}(\mathcal{O}(X)) = \sigma(\mathcal{O}(X)) \cap \text{Filt}(\mathcal{O}(X))$. The members of $\text{OFilt}(\mathcal{O}(X))$ are called open filters of X . For each $K \in \mathcal{K}(X)$, let $\Phi(K) = \{U \in \mathcal{O}(X) : K \subseteq U\}$. Then $\Phi(K) \in \text{OFilt}(\mathcal{O}(X))$ and $K = \bigcap \Phi(K)$. Obviously, $\Phi : \mathcal{K}(X) \rightarrow \text{OFilt}(\mathcal{O}(X))$, $K \mapsto \Phi(K)$, is an order embedding.

For a T_0 space X , $\mathcal{G} \subseteq 2^X$ and $W \subseteq X$, let $\diamond_{\mathcal{G}} W = \{G \in \mathcal{G} : G \cap W \neq \emptyset\}$ and $\square_{\mathcal{G}} W = \{G \in \mathcal{G} : G \subseteq W\}$. The symbols $\diamond_{\mathcal{G}} W$ and $\square_{\mathcal{G}} W$ will be simply written as $\diamond A$ and $\square A$ respectively if no ambiguity occur. The *lower Vietoris topology* on \mathcal{G} is the topology that has $\{\diamond U : U \in \mathcal{O}(X)\}$ as a subbase, and the resulting space is denoted by $P_H(\mathcal{G})$. The space $P_H(\mathcal{C}(X) \setminus \{\emptyset\})$, $P_H(X)$ for short, is called the *Hoare power space* or *lower space* of X (cf. [27]). Clearly, $P_H(X) = (\mathcal{C}(X) \setminus \{\emptyset\}, v(\mathcal{C}(X) \setminus \{\emptyset\}))$, and hence $P_H(X)$ is always sober (see [51, Corollary 4.10] or [42, Proposition 2.9]). The *upper Vietoris topology* on $\mathcal{K}(X)$ is the topology generated by $\{\square U : U \in \mathcal{O}(X)\}$ as a base, and the resulting space is called the *Smyth power space* or *upper space* of X and is denoted by $P_S(X)$ (cf. [8, 9, 27]).

Remark 2.6. Let X be a T_0 space.

1. If $\mathcal{S}_c(X) \subseteq \mathcal{G}$, then the specialization order on $P_H(\mathcal{G})$ is the set inclusion order, and the *canonical mapping* $\eta_X : X \rightarrow P_H(\mathcal{G})$, given by $\eta_X(x) = \overline{\{x\}}$, is a topological embedding (cf. [4, 7, 27]).
2. The space $X^s = P_H(\text{Irr}_c(X))$ with the canonical mapping $\eta_X : X \rightarrow X^s$ is the (*standard*) *sobrification* of X (cf. [4, 7]).

3. the specialization order on $P_S(X)$ is the Smyth order, that is, $\leq_{P_S(X)} = \sqsubseteq$;
4. the canonical mapping $\xi_X : X \rightarrow P_S(X)$, $x \mapsto \uparrow x$, is an order and topological embedding (cf. [8, 9, 27])

For a nonempty subset C of a T_0 space X , it is easy to see that C is compact iff $\uparrow C \in K(X)$. Furthermore, we have the following useful result (see, e.g., [10]).

Lemma 2.7. *Let X be a T_0 space and $C \in K(X)$. Then $C = \uparrow \min(C)$ and $\min(C)$ is compact.*

The Smyth power space construction defines a covariant functor. More precisely, we have the following.

Lemma 2.8. ([37]) *$P_S : \mathbf{Top}_0 \rightarrow \mathbf{Top}_0$ is a covariant functor, where for any $f : X \rightarrow Y$ in \mathbf{Top}_0 , $P_S(f) : P_S(X) \rightarrow P_S(Y)$ is defined by $P_S(f)(K) = \uparrow f(K)$ for all $K \in K(X)$.*

A T_0 space X is called *well-filtered* if for any open set U and any $K \in \mathcal{D}(K(X))$, $\cap K \subseteq U$ implies $K \subseteq U$ for some $K \in K$. The category of all well-filtered spaces and continuous mappings is denoted by \mathbf{Top}_w .

As in [54], a topological space X is *locally hypercompact* (resp., a *C-space*) if for each $x \in X$ and each open neighborhood U of x , there is $\uparrow F \in \mathbf{Fin} X$ such that $x \in \text{int } \uparrow F \subseteq \uparrow F \subseteq U$ (resp., there is $u \in X$ such that $x \in \text{int } \uparrow u \subseteq \uparrow u \subseteq U$). A set $K \subseteq X$ is called *supercompact* if for any family $\{U_i : i \in I\} \subseteq \mathcal{O}(X)$, $K \subseteq \cup_{i \in I} U_i$ implies $K \subseteq U$ for some $i \in I$. It is easy to verify that the supercompact saturated sets of X are exactly the sets $\uparrow x$ with $x \in X$ (see [9], Fact 2.2). It is well-known that X is a *C-space* iff $\mathcal{O}(X)$ is a *completely distributive* lattice (cf. [55]). A space X is called *core compact* if $(\mathcal{O}(X), \subseteq)$ is a *continuous lattice* (cf. [4]).

For a full subcategory \mathbf{K} of \mathbf{Top}_0 and an object X of \mathbf{K} , we will call X a \mathbf{K} -space. The category \mathbf{K} is said to be a *Keimel-Lawson category* if the following four properties are satisfied:

- (K₁) Homeomorphic copies of \mathbf{K} -spaces are \mathbf{K} -spaces.
- (K₂) $\mathbf{Sob} \subseteq \mathbf{K}$, that is, all sober spaces are \mathbf{K} -spaces.
- (K₃) In a sober space X , the intersection of any family of \mathbf{K} -subspaces of X is a \mathbf{K} -space.
- (K₄) Continuous maps $f : S \rightarrow T$ between sober spaces S and T are

\mathbf{K} -continuous, that is, for every \mathbf{K} -subspace K of T , the inverse image $f^{-1}(K)$ is a \mathbf{K} -subspace of S .

In what follows, \mathbf{K} always refers to a full subcategory of \mathbf{Top}_0 containing \mathbf{Sob} . The category \mathbf{K} is said to be *closed with respect to homeomorphisms* if \mathbf{K} has (K₁).

3. Well-filtered spaces and sober spaces

It is well-known that every sober space is well-filtered (see [14]) and every well-filtered space is a *d-space* (cf. [33, 42]). The Scott space of every continuous dcpo is sober (see [4]). Furthermore, the Scott space of every quasicontinuous domain is sober (see [6]). Johnstone [17] constructed the first dcpo whose Scott space is non-sober. Soon after, Isbell [15] gave a complete lattice whose Scott space is non-sober. The general problem in this line is whether each object in a classic class of dcpos or complete lattices has a sober Scott space.

Example 3.1. (Johnstone's dcpo) Let $\mathbb{J} = \mathbb{N} \times (\mathbb{N} \cup \{\infty\})$ with ordering defined by $(j, k) \leq (m, n)$ iff $j = m$ and $k \leq n$, or $n = \infty$ and $k \leq m$ (see Figure 1).

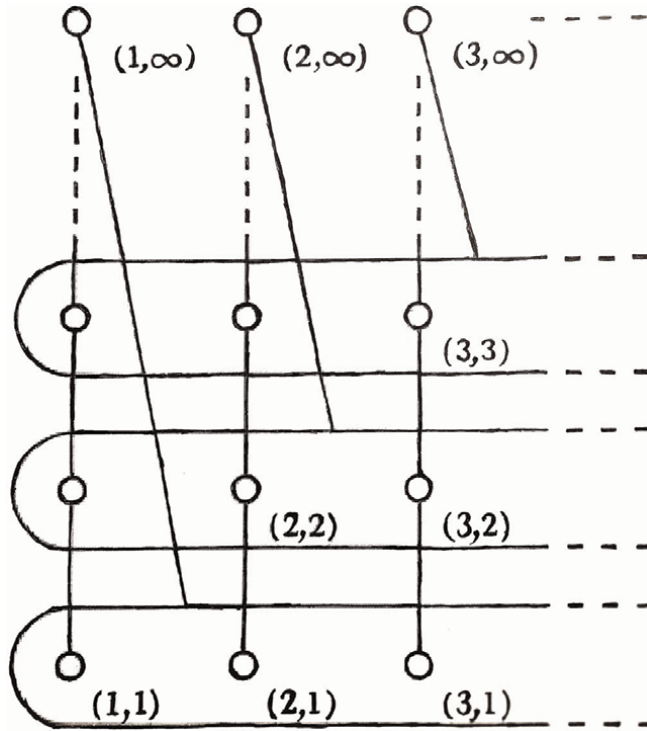


Figure 1.
 Johnstone's dcpo \mathbb{J} .

\mathbb{J} is a well-known dcpo constructed by Johnstone in [17]. Now, we show that \mathbb{J} is not well-filtered. Suppose, on the contrary, that there exists a well-filtered topology τ on \mathbb{J} which is compatible with the original order of \mathbb{J} . Clearly, $\bigcap_{n \in \mathbb{N}} (\uparrow(1, n) \cap \uparrow(2, 1)) = \emptyset$, and $\uparrow(1, n) \cap \uparrow(2, 1) = \{(m, \infty) : n \leq m\}$ is a compact saturated subset in $\Sigma\mathbb{J}$, and hence a compact saturated subset in (\mathbb{J}, τ) since $\tau \subseteq \sigma(\mathbb{J})$. By the well-filteredness of (\mathbb{J}, τ) , $\uparrow(1, n_0) \cap \uparrow(2, 1) = \emptyset$ for some $n_0 \in \mathbb{N}$, a contradiction. Thus, \mathbb{J} is not well-filtered. In particular, $\Sigma\mathbb{J}$ is not well-filtered (see [7], Exercise 8.3.9).

In 1992, Heckmann [8] asked whether every well-filtered dcpo is sober in its Scott topology. Kou [20] constructed the first dcpo whose Scott space is well-filtered but non-sober (see [16, 33, 52] for other different counterexamples) and hence gave a negative answer to Heckmann's question.

In [33], Xi and Lawson found out a condition under which a d -space is well-filtered.

Theorem 3.2. ([33]) *Let X be a d -space. If $\downarrow(A \cap K)$ is closed for any $A \in \mathcal{C}(X)$ and $K \in \mathcal{K}(X)$, then X is well-filtered.*

Corollary 3.3. ([33]) *For a dcpo P , if $\wedge P$ is compact (in particular, P is a complete lattice), then ΣP is well-filtered.*

It follows from Theorem 3.2 or Corollary 3.3 that Isbell's non-sober complete lattice is well-filtered. Note that Johnstone's dcpo is countable and the Isbell's non-sober complete lattice L is neither distributive nor countable. Thus in 1994, Abramsky and Jung asked whether there is a distributive complete lattice whose Scott space is non-sober (see [1], Exercises 7.3.19–6 or [18]). In [44], using Isbell's lattice, Xu, Xi, and Zhao gave a positive answer to this problem.

Theorem 3.4. ([44]) *Let L be the Isbell's non-sober lattice. Then $K(\Sigma L)$ is a spatial frame and the Scott space of $K(\Sigma L)$ is non-sober.*

For the lattice of closed subsets of ΣL , we have the following.

Proposition 3.5. ([25]) *Let L be the Isbell's lattice. Then ΣL is a retract of $\Sigma C(\Sigma L)$ and hence $\Sigma C(\Sigma L)$ is non-sober.*

In [56], Ern  showed that for the complete Boolean algebra B of all regular open subsets of the reals, the Scott space ΣB is not a topological join-semilattice (and hence the Scott topology $\sigma(B \times B)$ is strictly finer than the product topology $\sigma(B) \times \sigma(B)$). It is natural to wonder whether ΣB is sober. Thus Ern  asked the following.

Question 3.6. Let B be the complete Boolean algebra of all regular open subsets of the reals. Is the Scott space ΣB sober?

If the answer is negative, we would have a natural complete Boolean algebra whose Scott space is non-sober.

More generally, we pose the following question.

Question 3.7. ([47]) Is there a complete Boolean algebra B such that ΣB is not sober?

In 2019, in a talk [18] given at the National Institute of Education, Singapore, Achim Jung also asked whether there is a countable complete lattice whose Scott space is non-sober.

In [26], Miao, Xi, Li, and Zhao gave a negative answer to this question.

Theorem 3.8. ([26]) *There is a countably infinite complete lattice whose Scott space is non-sober.*

One of the most important results on sober spaces is the Hofmann-Mislove Theorem (see [14], Theorem 2.16) or [4], Theorem II-1.20 and Theorem II-1.21).

Theorem 3.9. (Hofmann-Mislove Theorem) *For a T_0 space X , the following conditions are equivalent:*

1. X is a sober space.
2. For any $\mathcal{F} \in \text{OFilt}(\mathcal{O}(X))$, there is a $K \in K(X) \cup \{\emptyset\}$ such that $\mathcal{F} = \Phi(K)$.
3. For any $\mathcal{F} \in \text{OFilt}(\mathcal{O}(X))$, $\mathcal{F} = \Phi(\cap \mathcal{F})$.

By Hofmann-Mislove Theorem, $\Phi : K(X) \cup \{\emptyset\} \rightarrow \text{OFilt}(\mathcal{O}(X))$ is an order isomorphism if and only if X is sober.

In [16], Jia asked whether every core compact well-filtered space is sober. This problem was independently answered in [21] and in [41, 42] by different methods.

Theorem 3.10. ([21, 41, 42]) *Every core compact well-filtered space is sober.*

In addition, the following result was obtained by Xu et al. in [41].

Theorem 3.11. ([41]) *Every first countable well-filtered T_0 space is sober.*

By ([20], Theorem 2.3), ([4], Theorem V-5.6), and Theorem 3.10, one get the following important result.

Theorem 3.12. *For a T_0 space X , the following conditions are equivalent:*

1. X is locally compact and sober.
2. X is locally compact and well-filtered.
3. X is core compact and sober.
4. X is core compact and well-filtered.

4. Rudin sets and well-filtered determined sets

Rudin's Lemma is a very useful tool in non-Hausdorff topology and plays a crucial role in domain theory (see [1, 4, 7, 8]). The following topological variant of Rudin's Lemma was given in [9].

Lemma 4.1. (Topological Rudin Lemma) *Let X be a topological space and \mathcal{A} an irreducible subset of the Smyth power space $P_S(X)$. Then every closed set $C \subseteq X$ that meets all members of \mathcal{A} contains a minimal irreducible closed subset A that still meets all members of \mathcal{A} .*

For a T_0 space X and $\mathcal{K} \subseteq \mathbf{K}(X)$, let $M(\mathcal{K}) = \{A \in \mathcal{C}(X) : K \cap A \neq \emptyset \text{ for all } K \in \mathcal{K}\}$ (that is, $\mathcal{A} \subseteq \diamond A$) and $m(\mathcal{K}) = \{A \in \mathcal{C}(X) : A \text{ is a minimal member of } M(\mathcal{K})\}$.

Definition 4.2. ([42]) A T_0 space X is called a *directed closure space*, DC space for short, if $\text{Irr}_c(X) = \mathcal{D}_c(X)$, that is, for each $A \in \text{Irr}_c(X)$, there exists a directed subset of X such that $A = \overline{D}$.

Remark 4.3. In [42], it was shown that closed subspaces, retracts and products of DC spaces are again DC spaces.

Based on the topological Rudin Lemma, the concept of Rudin spaces can be defined (see [29, 42]).

Definition 4.4. Let X be a T_0 space and A a nonempty subset of X .

1. A is said to have the *Rudin property*, if there exists a filtered family $\mathcal{K} \subseteq \mathbf{K}(X)$ such that $\overline{A} \in m(\mathcal{K})$ (that is, \overline{A} is a minimal closed set that intersects all members of \mathcal{K}). Let $\text{RD}(X) = \{A \in \mathcal{C}(X) : A \text{ has Rudin property}\}$. The sets in $\text{RD}(X)$ will also be called *Rudin sets*.
2. X is called a *Rudin space*, *RD space* for short, if $\text{Irr}_c(X) = \text{RD}(X)$, that is, every irreducible closed set of X is a Rudin set.

It was proved in [29] that the closed subspaces, retracts and product of Rudin spaces are again Rudin spaces.

Definition 4.5. ([36]) Let \mathbf{K} be a full category of \mathbf{Top}_0 containing **Sob** and X a T_0 space.

1. A subset A of X is said to be **K-determined**, provided for any continuous mapping $f : X \rightarrow Y$ to a **K**-space Y , there exists a unique $y_A \in Y$ such that $\overline{f(A)} = \overline{\{y_A\}}$. Denote by $\mathbf{K}(X)$ the set of all closed **K**-determined sets of X .
2. The space X is said to be a **K-determined** space, if $\text{Irr}_c(X) = \mathbf{K}(X)$ or, equivalently, all irreducible closed sets of X are **K**-sets.

For simplicity, let $d(X) = \mathbf{Top}_d(X)$ and $\text{WD}(X) = \mathbf{Top}_w(X)$. The sets in $\text{WD}(X)$ (resp., $d(X)$) are called **WD sets** (resp., *d-determined sets*). The space X is called a *well-filtered determined* space, shortly a **WD space**, if all irreducible closed subsets of X are **WD sets**, that is, $\text{Irr}_c(X) = \text{WD}(X)$ (see [36, 42]). The space X is called a *d-determined* space if $\text{Irr}_c(X) = d(X)$.

Proposition 4.6. ([36, 42]) *Let \mathbf{K} be a full category of \mathbf{Top}_0 containing **Sob** and X a T_0 space. Then*

1. $S_c(X) \subseteq \mathbf{K}(X) \subseteq \mathbf{Sob}(X) = \text{Irr}_c(X)$.

$$2. S_c(X) \subseteq \mathcal{D}_c(X) \subseteq \text{RD}(X) \subseteq \text{WD}(X) \subseteq \text{Irr}_c(X) \text{ and } \mathcal{D}_c(X) \subseteq d(X) \subseteq \text{WD}(X).$$

$$3. \text{Sober} \Rightarrow \text{DC} \Rightarrow \text{RD} \Rightarrow \text{WD}.$$

4. Every sober space is \mathbf{K} -determined.

From Proposition 4.6, we know that Rudin spaces lie between WD spaces and DC spaces. Also, DC spaces lie between Rudin spaces and sober spaces.

In ([22], Example 4.15), a T_0 space X was constructed in which some well-filtered determined sets are not Rudin sets and hence gave a negative answer to a question posed by Xu and Zhao in [46]: Does $\text{RD}(X) = \text{WD}(X)$ hold for every T_0 space X ? It is not difficult to check that the space X is a WD space but not a Rudin space. Therefore, Example 4.15 in [22] also gave a negative answer to another related question raised by Xu et al. in [42]: Is every well-filtered determined space a Rudin space?

Proposition 4.7. ([36]) *Suppose that \mathbf{K} is adequate and closed with respect to homeomorphisms. Then a T_0 space X is a \mathbf{K} -space if and only if $\mathbf{K}(X) = S_c(X)$.*

Using Rudin sets and WD sets, one can obtain some characterizations of well-filtered spaces and sober spaces.

Proposition 4.8. ([42]) *For a T_0 space X , the following conditions are equivalent:*

1. X is well-filtered.
2. $\text{RD}(X) = S_c(X)$.
3. $\text{WD}(X) = S_c(X)$.

Proposition 4.9. ([42]) *For a T_0 space X , the following conditions are equivalent:*

1. X is sober.
2. X is a DC d -space.
3. X is a well-filtered DC space.
4. X is a well-filtered Rudin space.
5. X is a well-filtered WD space.

Lemma 4.10. ([54]) *If X is a locally hypercompact T_0 space and $A \in \text{Irr}(X)$, then there is a directed subset $D \subseteq \downarrow A$ such that $\overline{A} = \overline{D}$ for. Therefore, X is a DC space.*

Theorem 4.11. ([42]) *Let X be a T_0 space.*

1. If X is locally compact, then it is a Rudin space.
2. If X is core compact, then it is a WD space.

Figure 2 shows some relationships among some types of spaces.

Question 4.12. ([42]) *Is every core compact T_0 space a Rudin space?*

Theorem 4.13. ([38]) *Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing \mathbf{Sob} and $X = \prod_{i \in I} X_i$ the product of a family $\{X_i : i \in I\}$ of T_0 spaces. For $A \in \text{Irr}(X)$, the following conditions are equivalent:*

1. A is a \mathbf{K} -determined set.
2. $p_i(A)$ is a \mathbf{K} -determined set for each $i \in I$.

From Lemma 2.5 and Theorem 4.13, we deduce the following.

Corollary 4.14. *Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing **Sob** and $X = \prod_{i \in I} X_i$ the product of a family $\{X_i : i \in I\}$ of T_0 spaces. If $A \in \mathbf{K}(X)$, then $A = \prod_{i \in I} p_i(X_i)$, and $p_i(A) \in \mathbf{K}(X_i)$ for all $i \in I$.*

Theorem 4.15. ([38]) *Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing **Sob** and $X = \prod_{i \in I} X_i$ the product of a family $\{X_i : i \in I\}$ of T_0 spaces. Then the following conditions are equivalent:*

1. X is a \mathbf{K} -determined space.
2. For each $i \in I$, X_i is a \mathbf{K} -determined space.

Let $\mathbf{K} = \mathbf{Top}_d$ or $\mathbf{K} = \mathbf{Top}_w$. Then by Theorems 4.13 and 4.15, we deduce the following two corollaries.

Corollary 4.16. ([38]) *Let $\{X_i : i \in I\}$ be a family of T_0 spaces and $X = \prod_{i \in I} X_i$ the product space. For $A \in \text{Irr}(X)$, the following conditions are equivalent:*

1. A is d -determined (resp., well-filtered determined).
2. $p_i(A)$ is d -determined (resp., well-filtered determined) for each $i \in I$.

Corollary 4.17. ([38]) *For a family $\{X_i : i \in I\}$ of T_0 spaces, the following two conditions are equivalent:*

1. The product space $\prod_{i \in I} X_i$ is a d -determined space (resp., a $\text{mathsf{fWD}}$ space).
2. For each $i \in I$, X_i is a d -determined space (resp., a $\text{mathsf{fWD}}$ space).

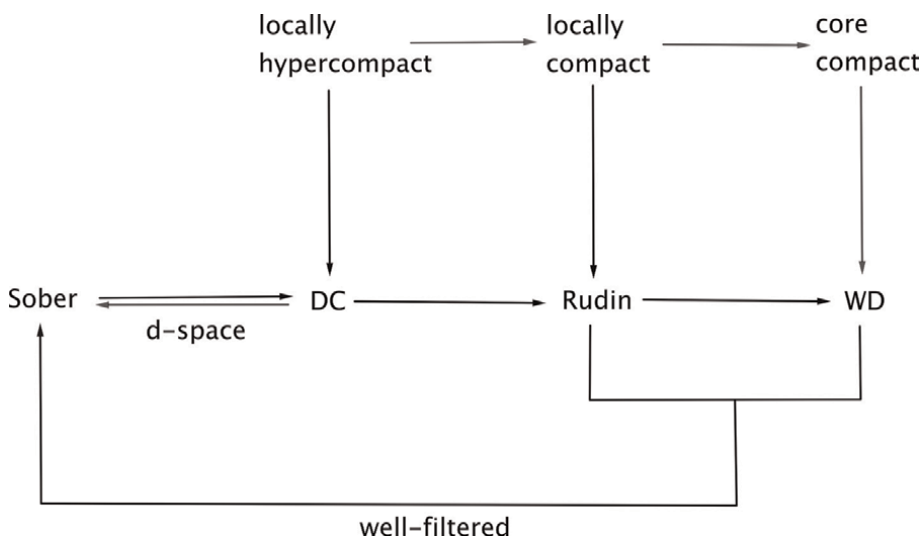


Figure 2.
 Certain relations among some kinds of spaces.

5. K-reflections of T_0 spaces

Definition 5.1. ([19, 36]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_0 containing **Sob** and X a T_0 space. A \mathbf{K} -reflection of X is a pair $\langle \tilde{X}, \mu \rangle$ consisting of a \mathbf{K} -space \tilde{X} and a continuous mapping $\mu : X \rightarrow \tilde{X}$ satisfying that for any continuous mapping $f : X \rightarrow Y$ to a \mathbf{K} -space, there exists a unique continuous mapping $f^* : \tilde{X} \rightarrow Y$ such that $f^* \circ \mu = f$, that is, the following diagram commutes.

By a standard argument, \mathbf{K} -reflections, if they exist, are unique up to homeomorphism. We shall use X^k to denote the space of the \mathbf{K} -reflection of X if it exists. The space of **Sob**-reflection of X is the sobrification X^s of X . The space of \mathbf{Top}_d -reflection (resp., \mathbf{Top}_w -reflection) of X is denoted by X^d (resp., X^w).

It is well-known that **Sob** is reflective in \mathbf{Top}_0 (see [4, 7]). Using d -closures, Wyler [32] proved that \mathbf{Top}_d is reflective in \mathbf{Top}_0 (see also [2, 3, 48]). In [19], it was proved by Wyler's method that every Keimel-Lawson category \mathbf{K} is reflective in \mathbf{Top}_0 .

For quite a long time, it is not known whether \mathbf{Top}_w is reflective in \mathbf{Top}_0 . Recently, this problem was answered positively and independently in [22, 29, 31, 42]. In [36], for a full subcategory \mathbf{K} of \mathbf{Top}_0 containing **Sob**, a direct approach to \mathbf{K} -reflections of T_0 spaces were provided.

By Proposition 4.6, $\{\circ_{\mathbf{K}(X)}U : U \in \mathcal{O}(X)\}$ is a topology on $\mathbf{K}(X)$. In the following, let $\eta_X^k : X \rightarrow P_H(\mathbf{K}(X))$, $\eta_X^k(x) = \overline{\{x\}}$, be the canonical mapping from X to $P_H(\mathbf{K}(X))$. It is straightforward to verify that the canonical mapping $\eta_X^k : X \rightarrow P_H(\mathbf{K}(X))$ is a topological embedding.

For the \mathbf{K} -reflections of T_0 spaces, the following lemma is crucial.

Lemma 5.2. ([36]) Let X be a T_0 space and $f : X \rightarrow Y$ a continuous mapping from X to a well-filtered space Y . Then there exists a unique continuous mapping $f^* : P_H(\mathbf{K}(X)) \rightarrow Y$ such that $f^* \circ \eta_X^k = f$, that is, the following diagram commutes.

Theorem 5.3. ([36]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_0 containing **Sob** and Let X a T_0 space. If $P_H(\mathbf{K}(X))$ is a \mathbf{K} -space, then the pair $\langle X^k = P_H(\mathbf{K}(X)), \eta_X^k \rangle$, where $\eta_X^k : X \rightarrow X^k$, $x \mapsto \overline{\{x\}}$, is the \mathbf{K} -reflection of X .

For a full subcategory \mathbf{K} of \mathbf{Top}_0 containing **Sob**, \mathbf{K} is called *adequate* if $P_H(\mathbf{K}(X))$ is a \mathbf{K} -space for any T_0 space X .

By Theorem 5.3, we get the following.

Corollary 5.4. ([36]) If \mathbf{K} is adequate, then \mathbf{K} is reflective in \mathbf{Top}_0 . For a T_0 space X , $X^k = P_H(\mathbf{K}(X))$ with canonical mapping $\eta_X : X \rightarrow X^k$ is the \mathbf{K} -reflection of X .

Theorem 5.5. ([36]) If \mathbf{K} is adequate, then for any T_0 spaces X, Y and any continuous mapping $f : X \rightarrow Y$, there exists a unique continuous mapping $f^k : X^k \rightarrow Y^k$ such that $f^k \circ \eta_X^k = \eta_Y^k \circ f$, that is, the following diagram commutes.

For each $A \in \mathbf{K}(X)$, $f^k(A) = \overline{f(A)}$. Therefore, a functor $K : \mathbf{Top}_0 \rightarrow \mathbf{K}$ is defined, which is the left adjoint to the inclusion functor $I : \mathbf{K} \rightarrow \mathbf{Top}_0$.

Theorem 5.6. ([36]) If $\mathbf{K} \in \{\mathbf{Sob}, \mathbf{Top}_d, \mathbf{Top}_w\}$ or \mathbf{K} is a Keimel-Lawson category, then \mathbf{K} is adequate and hence reflective in \mathbf{Top}_0 . For a T_0 space X , $X^k = P_H(\mathbf{K}(X))$ with the canonical mapping $\eta_X : X \rightarrow X^k$ is the \mathbf{K} -reflection of X .

Corollary 5.7. ([36]) \mathbf{Top}_d is adequate and hence reflective in \mathbf{Top}_0 . For a T_0 space X , $X^d = P_H(d(X))$ with the canonical mapping $\eta_X : X \rightarrow X^d$ is the d -reflection of X .

Corollary 5.8. ([36]) \mathbf{Top}_w is adequate and hence reflective in \mathbf{Top}_0 . For a T_0 space X , $X^w = P_H(\mathbf{WD}(X))$ with the canonical mapping $\eta_X : X \rightarrow X^w$ is the well-filtered reflection of X .

6. K-reflections of Scott spaces

In this section, some necessary and sufficient conditions are given for the \mathbf{K} -reflection of a T_0 space (especially, a Scott space) to be a Scott space.

Theorem 6.1. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing \mathbf{Sob} which is adequate and closed with respect to homeomorphisms. For a T_0 space X , consider the following three conditions:

1. $\Sigma \mathbf{K}(X)$ is a \mathbf{K} -space.
2. $\eta_X^\sigma : X \rightarrow \Sigma \mathbf{K}(X)$, $\eta_X^\sigma(x) = \overline{\{x\}}$, is continuous.
3. The \mathbf{K} -reflection X^k of X is a Scott space.

Then (1) + (2) \Rightarrow (3), and (3) \Rightarrow (1). Moreover, when Conditions (1) and (2) hold, the Scott space $\Sigma \mathbf{K}(X)$ with the canonical mapping $\eta_X^\sigma : X \rightarrow \Sigma \mathbf{K}(X)$, $\eta_X^\sigma(x) = \overline{\{x\}}$, is a \mathbf{K} -reflection of X .

In particular, we have the following result for the Scott space of a poset.

Theorem 6.2. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing \mathbf{Sob} which is adequate and closed with respect to homeomorphisms. Then for any poset P , the following two conditions are equivalent:

1. $\Sigma \mathbf{K}(\Sigma P)$ is a \mathbf{K} -space.
2. The \mathbf{K} -reflection $(\Sigma P)^k$ of ΣP is a Scott space.

Moreover, when Condition (1) holds, the Scott space $\Sigma \mathbf{K}(\Sigma P)$ with the canonical mapping $\eta_P^\sigma : \Sigma P \rightarrow \Sigma \mathbf{K}(\Sigma P)$, $\eta_P^\sigma(x) = \text{cl}_{\sigma(P)}\{x\} = \downarrow x$, is a \mathbf{K} -reflection of ΣP .

For the case of $\mathbf{K} = \mathbf{Top}_d$, as an immediate corollary of Theorems 6.1 and 6.2), we get the following Corollary.

Corollary 6.3. ([19, 36]) Let X be a T_0 space and P a poset.

1. If $\xi_X : X \rightarrow \Sigma d(X)$, $\xi_X(x) = \overline{\{x\}}$, is continuous, then the d -reflection X^d of X is the Scott space $\Sigma d(X)$.
2. The poset $d(\Sigma P)$ is a dcpo and the d -reflection $(\Sigma P)^d$ of ΣP is the Scott space $\Sigma d(\Sigma P)$ with the canonical mapping $\eta_P : \Sigma P \rightarrow \Sigma d(X)$, given by $\eta_P(x) = \text{cl}_{\sigma(P)}\{x\}$ for each $x \in X$.

When $\mathbf{K} = \mathbf{Top}_w$ in Theorems 6.1 and 6.2, we get the following two corollaries.

Corollary 6.4. ([39]) For a T_0 space X , consider the following three conditions:

1. $\Sigma \mathbf{WD}(X)$ is a well-filtered space.

2. $\eta_X^w : X \rightarrow \Sigma \text{WD}(X)$, $\eta_X^w(x) = \overline{\{x\}}$, is continuous.

3. The well-filtered reflection X^w of X is a Scott space.

Then (1) + (2) \Rightarrow (3), and (3) \Rightarrow (1). Moreover, when Conditions (1) and (2) hold, the Scott space $\Sigma \text{WD}(X)$ with the canonical mapping $\eta_X^w : X \rightarrow \Sigma \text{WD}(X)$, $\eta_X^w(x) = \overline{\{x\}}$, is a well-filtered reflection of X .

Corollary 6.5. ([39]) For any poset P , the following two conditions are equivalent:

1. $\Sigma \text{WD}(\Sigma P)$ is a well-filtered space.

2. The well-filtered reflection $(\Sigma P)^w$ of ΣP is a Scott space.

Moreover, when Condition (1) holds, the Scott space $\Sigma \text{WD}(\Sigma P)$ with the canonical mapping $\eta_P^w : \Sigma P \rightarrow \Sigma \text{WD}(\Sigma P)$, $\eta_P^w(x) = \text{cl}_{\sigma(P)}\{x\} = \downarrow x$, is a well-filtered reflection of ΣP .

Theorem 6.6. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing **Sob** which is adequate and closed with respect to homeomorphisms. Suppose that X is a T_0 space for which $\text{Irr}_c(X) = \{\overline{\{x\}} : x \in X\} \cup \{X\}$ and X is not a \mathbf{K} -space. Consider the following three conditions:

1. $\Sigma (\Omega X)_T$ is a \mathbf{K} -space.

2. $\zeta_X^\sigma : X \rightarrow \Sigma (\Omega X)_T$, $\zeta_X^\sigma(x) = x$, is continuous.

3. The \mathbf{K} -reflection X^k of X is a Scott space.

Then (1) + (2) \Rightarrow (3), and (3) \Rightarrow (1). Moreover, when Conditions (1) and (2) hold, the Scott space $\Sigma (\Omega X)_T$ with the canonical mapping $\zeta_X^\sigma : X \rightarrow \Sigma (\Omega X)_T$, $\zeta_X^\sigma(x) = x$, is a \mathbf{K} -reflection of X .

Definition 6.7. ([39]) For a poset P , let $P_T = P \cup \{T\}$ ($T \notin P$) denote the poset obtained from P by adjoining the largest element T (whether P has one or not).

Clearly, the order on P_T is as follows: $x \leq y$ iff $x \leq y$ in P or $y = T$. The element T is the largest element of P_T (even P has the largest element).

It is straightforward to verify the following result.

Lemma 6.8. ([39]) Let P be a dcpo. Then

1. P_T is a dcpo and in P_T we have $T \ll T$, i.e., $\uparrow T = \{T\} \in \sigma(P_T)$,

2. $\zeta_P : \Sigma P \rightarrow \Sigma P_T$, $x \mapsto x$, is continuous.

By Lemma 6.8 and Theorem 6.6, we get the following.

Corollary 6.9. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing **Sob** which is adequate and closed with respect to homeomorphisms. Suppose that P is a poset for which $\text{Irr}_c(\Sigma P) = \{\overline{\{x\}} : x \in P\} \cup \{P\}$ and ΣP is not a \mathbf{K} -space. Then the following two conditions are equivalent:

1. ΣP_T is a \mathbf{K} -space.

2. The \mathbf{K} -reflection $(\Sigma P)^k$ of ΣP is a Scott space.

Moreover, when Condition (1) holds, the Scott space ΣP_T with the embedding $i_P : \Sigma P \rightarrow \Sigma P_T$, $i_P(x) = x$, is a \mathbf{K} -reflection of X .

Now we give some examples and counterexamples related to the \mathbf{K} -reflections of T_0 spaces (esp., Scott spaces).

Example 6.10. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing **Sob** which is adequate and closed with respect to homeomorphisms. Since \mathbb{N} is not a dcpo, $\Sigma \mathbb{N}$ is not a d -space and hence not a \mathbf{K} -space. Clearly, $\text{Irr}_c(\Sigma \mathbb{N}) = \{\overline{\{n\}} = \downarrow n : n \in \mathbb{N}\} \cup \{\mathbb{N}\}$. As \mathbb{N}_T is an algebraic lattice, by ([4], Proposition III-3.7), $\Sigma \mathbb{N}_T$ is a sober space and hence a \mathbf{K} -space. By Corollary 6.9, the \mathbf{K} -reflection of $\Sigma \mathbb{N}$ is a Scott space. More precisely, $\Sigma \mathbb{N}_T$ with the embedding $i_{\mathbb{N}} : \Sigma \mathbb{N} \rightarrow \Sigma \mathbb{N}_T$, $i_{\mathbb{N}}(n) = n$, is a \mathbf{K} -reflection of $\Sigma \mathbb{N}$.

Example 6.11. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing **Sob** which is adequate and closed with respect to homeomorphisms and $P = \mathbb{N} \cup \{a, b\}$. Define a partial order \leq on P as follows:

- i. $n < n + 1$ for each $n \in \mathbb{N}$,
- ii. $n < a$ and $n < b$ for all $n \in \mathbb{N}$, and
- iii. a and b are incomparable.

Then $\max(P) = \{a, b\}$ and P is not a dcpo since the chain \mathbb{N} does not have a least upper bound in P . So ΣP is not a d -space and hence not a \mathbf{K} -space. It is easy to verify that $\text{Irr}_c(\Sigma P) = \{\downarrow x : x \in P\} \cup \{\mathbb{N}\}$. Hence by Propositions 4.6 and 4.7, $\mathbf{K}(\Sigma P) = \text{Irr}_c(\Sigma P) = \{\downarrow x : x \in P\} \cup \{\mathbb{N}\}$. Now we show that $\Sigma \mathbf{K}(\Sigma P)$ is sober. Let $Q = \mathbb{N} \cup \{a, b, c\}$. Define a partial order \leq_Q on Q as follows:

- a. for $x, y \in P$, $x \leq_Q y$ iff $x \leq_P y$ in P ,
- b. $n <_Q c$ for all $n \in \mathbb{N}$, and
- c. $c <_Q a$ and $c <_Q b$.

Clearly, Q is an algebraic domain and $K(Q) = \mathbb{N} \cup \{a, b\}$. Define a mapping $\psi : \mathbf{K}(\Sigma P) \rightarrow Q$ by

$$\psi(x) = \begin{cases} n, & x = \uparrow n \ (n \in \mathbb{N}), \\ c, & x = \mathbb{N}, \\ a, & x = \{a\}, \\ b, & x = \{b\}. \end{cases}$$

It is straightforward to verify that ψ is a poset isomorphism, and hence induces a homeomorphism from $\Sigma \mathbf{K}(\Sigma P)$ to ΣQ . Clearly, Q is a dcpo, $K(Q) = \mathbb{N} \cup \{a, b\}$ and $c = \vee_Q \mathbb{N}$, whence Q is an algebraic domain. By ([4], Proposition III-3.7), ΣQ is sober, and consequently, $\Sigma \mathbf{K}(\Sigma P)$ is a sober space and hence a \mathbf{K} -space. It follows from Theorem 6.2 that the \mathbf{K} -reflection of ΣP is a Scott space. More precisely, ΣQ with the embedding $i_P = \psi \circ \eta_P^c : \Sigma P \rightarrow \Sigma Q$, $i_P(x) = x$, is a \mathbf{K} -reflection of ΣP .

The following two examples show that for a T_0 space X , Condition (1) of Theorem 6.1 is only a necessary condition but not a sufficient condition for the \mathbf{K} -reflection of X to be a Scott space.

Example 6.12. Let \mathbf{K} be a full subcategory of \mathbf{Top}_w containing \mathbf{Sob} which is adequate and closed with respect to homeomorphisms. Let X be a countably infinite set and endow X with the co-finite topology (having the finite sets as closed sets). The resulting space is denoted by X_{cof} . Then

- a. $\mathcal{C}(X_{cof}) = \{\emptyset, X\} \cup X^{(<\omega)}$, X_{cof} is T_1 and hence a d -space.
- b. $\mathbf{K}(X_{cof}) = 2^X \setminus \{\emptyset\}$.
- c. X_{cof} is locally compact and first-countable.
- d. X_{cof} is not well-filtered and hence not a \mathbf{K} -space.

Let $\mathcal{K} = \{X \setminus F : F \in X^{(<\omega)}\}$. Then \mathcal{K} is a filtered family of saturated compact subsets of X_{cof} and $\cap \mathcal{K} = \emptyset$, but $X \setminus F \neq \emptyset$ for every $F \in X^{(<\omega)}$. Thus X_{cof} is not well-filtered.

- e. $\mathbf{K}(X_{cof}) = \text{Irr}_c(X_{cof}) = \{\{x\} : x \in X\} \cup \{X\}$.

It is easy to see that $\text{Irr}_c(X_{cof}) = \{\{x\} : x \in X\} \cup \{X\}$. By Propositions 4.6 and 4.7, $\mathbf{K}(X_{cof}) = \text{Irr}_c(X_{cof}) = \{\{x\} : x \in X\} \cup \{X\}$.

- f. $\Sigma \mathbf{K}(X_{cof})$ is sober and hence a \mathbf{K} -space.

Clearly, $\mathbf{K}(X_{cof})$ (with the order of set inclusion) is a Noetherian dcpo, that is, it satisfies the *ascending chain condition*: every ascending chain has a greatest member. Whence $\mathbf{K}(X_{cof})$ is an algebraic domain. By ([4], Proposition III-3.7), $\Sigma \mathbf{K}(X_{cof})$ is sober, whence it is a \mathbf{K} -space.

- g. $\eta_{X_{cof}}^\sigma : X_{cof} \rightarrow \Sigma \text{Irr}_c(X_{cof})$, $x \mapsto \{x\}$, is not continuous.

Let $C \notin \mathcal{O}(X_{cof})$. Then $\{\{x\} : x \in C\} \cup \{X\} \in \sigma(\text{Irr}_c(X_{cof}))$, but

$(\eta_{X_{cof}}^\sigma)^{-1}(\{\{x\} : x \in C\} \cup \{X\}) = C \notin \mathcal{O}(X_{cof})$, proving that $\eta_{X_{cof}}^\sigma : X_{cof} \rightarrow \Sigma \text{Irr}_c(X_{cof})$ is not continuous.

- h. The \mathbf{K} -reflection of X_{cof} is not a Scott space. In particular, the well-filtered reflection of X_{cof} is not a Scott space and the sobrification of X_{cof} is also not a Scott space.

Assume, on the contrary, that the \mathbf{K} -reflection of X_{cof} is a Scott space. Then there is a poset P such that $(X_{cof})^k = P_H(\mathbf{K}(X_{cof}))$ is homeomorphic to ΣP , whence by (e), $\text{Irr}_c(X_{cof}) = \mathbf{K}(X_{cof}) (= \Omega P_H(\text{Irr}_c(X_{cof})))$ and $P (= \Omega \Sigma P)$ are isomorphic. It follows that $\Sigma \text{Irr}_c(X_{cof})$ and ΣP are homeomorphic, and consequently, $P_H(\text{Irr}_c(X_{cof})) \cong \Sigma \text{Irr}_c(X_{cof})$. Therefore, $\mathcal{O}(P_H(\text{Irr}_c(X_{cof}))) \cong \sigma(\text{Irr}_c(X_{cof}))$, and hence $2^\omega = |\sigma(\text{Irr}_c(X_{cof}))| = |\mathcal{O}(P_H(\text{Irr}_c(X_{cof})))| \leq |\mathcal{O}(X_{cof})| = |X^{(<\omega)}| = \omega$, which is a contradiction by Cantor's Theorem (see [57], III-2.13 Cantor's Theorem). So the \mathbf{K} -reflection of X_{cof} is not a Scott space.

Example 6.13. ([39]) Let $X = 2^{\mathbb{N}}$ (the set of all subsets of \mathbb{N}) and endow X with the *co-countable topology* (having the countable sets as closed sets). The resulting space is denoted by X_{coc} . Then

- a. $|X| = \mathfrak{c} = 2^{\omega}$ (where $\mathfrak{c} = |\mathbb{R}|$ and \mathbb{R} is the set of all reals) and X is an uncountably infinite set.
- b. X_{coc} is T_1 and $\mathcal{C}(X_{coc}) = \{\emptyset, X\} \cup X^{(\leq \omega)}$.
- c. $K(X_{coc}) = X^{(< \omega)} \setminus \{\emptyset\}$.
- d. X_{coc} is well-filtered.
- e. $\text{Irr}_c(X_{coc}) = \{\{x\} : x \in X\} \cup \{X\}$.
- f. $\Sigma \text{Irr}_c(X_{coc})$ is sober.

Let $P = \{\{x\} : x \in X\} \cup \{X\}$ with the order of set inclusion. It is easy to see that P is a Noetherian dcpo and hence ΣP is sober by ([4], Proposition III-3.7). Clearly, $\sigma(P) = \gamma(P)$ and hence $|\sigma(P)| = |\gamma(P)| = 2^{\mathfrak{c}}$.

- g. $\eta_{X_{coc}}^{\sigma} : X_{coc} \rightarrow \Sigma \text{Irr}_c(X_{coc})$, $x \mapsto \{x\}$, is not continuous.

Let C be any non-countable proper subset of X , that is, $C \notin \mathcal{C}(X_{coc})$. Then $\{\{x\} : x \in C\} \in \mathcal{C}(\Sigma(\text{Irr}_c(X_{coc})))$, but $(\eta_{X_{coc}}^{\sigma})^{-1}(\{\{x\} : x \in C\}) = C \notin \mathcal{C}(X_{coc})$, proving that $\eta_{X_{coc}}^{\sigma} : X_{coc} \rightarrow \Sigma \text{Irr}_c(X_{coc})$ is not continuous.

- h. The sobrification of X_{coc} is not a Scott space.

Assume, on the contrary, that the sobrification of X_{coc} is a Scott space. Then there is a poset P such that $(X_{coc})^s = P_H(\text{Irr}_c(X_{coc}))$ is homeomorphic to ΣP , whence $\text{Irr}_c(X_{coc}) (= \Omega P_H(\text{Irr}_c(X_{coc})))$ and $P (= \Omega \Sigma P)$ are isomorphic. It follows that $\Sigma \text{Irr}_c(X_{coc})$ and ΣP are homeomorphic, and consequently, $P_H(\text{Irr}_c(X_{coc})) \cong \Sigma \text{Irr}_c(X_{coc})$. Therefore, $\mathcal{O}(P_H(\text{Irr}_c(X_{coc}))) \cong \sigma(\text{Irr}_c(X_{coc}))$, and hence $2^{\mathfrak{c}} = |\sigma(\text{Irr}_c(X_{coc}))| = |\mathcal{O}(P_H(\text{Irr}_c(X_{coc})))| \leq |\mathcal{O}(X_{coc})| = (2^{\omega})^{\omega} = 2^{\omega \cdot \omega} = 2^{\omega} = \mathfrak{c}$ (see [57], III-3.23 Corollary and III-3.29 Proposition), which is a contradiction by Cantor's Theorem. So the sobrification of X_{coc} is not a Scott space.

Proposition 6.14. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_w containing **Sob** which is adequate and closed with respect to homeomorphisms. Then the \mathbf{K} -reflection of Johnstone's space $\Sigma \mathbb{J}$ is not a Scott space. In particular, neither the sobrification nor the well-filtered reflection of $\Sigma \mathbb{J}$ is a Scott space.

7. K-reflections of product spaces

Theorem 7.1. ([38]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing **Sob**. If \mathbf{K} is adequate, then for any family $\{X_i : i \in I\}$ of T_0 spaces, $(\prod_{i \in I} X_i)^k = \prod_{i \in I} X_i^k$ (up to homeomorphism).

Corollary 7.2. ([11, 12]) *For any family $\{X_i : i \in I\}$ of T_0 spaces, $(\prod_{i \in I} X_i)^s = \prod_{i \in I} X_i^s$ (up to homeomorphism).*

By Proposition 4.6, **Sob** is adequate, and hence by Theorem 7.1, we have the following result.

Corollary 7.3. ([11, 12]) *For a family $\{X_i : i \in I\}$ of T_0 spaces, $(\prod_{i \in I} X_i)^s = \prod_{i \in I} X_i^s$ (up to homeomorphism).*

By Corollary 5.7 and Theorem 7.1, we have the following corollary.

Corollary 7.4. ([38]) *For a family $\{X_i : i \in I\}$ of T_0 spaces, $(\prod_{i \in I} X_i)^d = \prod_{i \in I} X_i^d$ (up to homeomorphism).*

Proposition 7.5. ([38]) *Let $\{P_i : i \in I\}$ be a family of posets. If the Scott topology of the product $\prod_{i \in I} P_i$ is the product of the Scott topologies of the factors, then the d -reflection $(\prod_{i \in I} \Sigma P_i)^d$ of $\prod_{i \in I} \Sigma P_i$ is the Scott space $\Sigma d(\Sigma \prod_{i \in I} P_i)$.*

By Corollary 5.8 and Theorem 7.1, we deduce the following result.

Corollary 7.6. ([38]) *For a family $\{X_i : i \in I\}$ of T_0 spaces, $(\prod_{i \in I} X_i)^w = \prod_{i \in I} X_i^w$ (up to homeomorphism).*

By Theorems 5.6 and 7.1, we have the following corollary.

Corollary 7.7. ([38]) *Let \mathbf{K} be a Keimel-Lawson category. If \mathbf{K} is a full subcategory of \mathbf{Top}_d , then for any family $\{X_i : i \in I\}$ of T_0 spaces, $(\prod_{i \in I} X_i)^k = \prod_{i \in I} X_i^k$ (up to homeomorphism).*

Proposition 7.8. ([38]) *Let $H : \mathbf{Top}_0 \rightarrow \mathbf{Set}$ be an R -subset system and $X = \prod_{i \in I} X_i$ the product space of a family $\{X_i : i \in I\}$ of T_0 spaces. For $A \in \text{Irr}(X)$, the following two conditions are equivalent:*

1. A is H -sober determined (resp., super H -sober determined).
2. $p_i(A)$ is H -sober determined (resp., super H -sober determined) for each $i \in I$.

Proposition 7.9. *Let $H : \mathbf{Top}_0 \rightarrow \mathbf{Set}$ be an R -subset system and $\{X_i : i \in I\}$ a family of T_0 spaces. Then $(\prod_{i \in I} X_i)^h = \prod_{i \in I} X_i^h$ (up to homeomorphism).*

Proposition 7.10. *Let $H : \mathbf{Top}_0 \rightarrow \mathbf{Set}$ be an R -subset system having property M and $\{X_i : i \in I\}$ a family of T_0 spaces. Then $(\prod_{i \in I} X_i)^H = \prod_{i \in I} X_i^H$ (up to homeomorphism).*

For the concepts and notations used in the above three propositions, please see [41].

From Theorem 7.1, we deduce the following.

Corollary 7.11. ([36]) *Suppose that \mathbf{K} is adequate and closed with respect to homeomorphisms. Then for any family $\{X_i : i \in I\}$ of T_0 spaces, the following two conditions are equivalent:*

1. The product space $\prod_{i \in I} X_i$ is a \mathbf{K} -space.
2. For each $i \in I$, X_i is a \mathbf{K} -space.

By Theorem 5.6 and Corollary 7.11, we get the following corollary.

Corollary 7.12. *If $\mathbf{K} \in \{\mathbf{Sob}, \mathbf{Top}_d, \mathbf{Top}_w\}$ or \mathbf{K} is a Keimel-Lawson category, then for any family $\{X_i : i \in I\}$ of T_0 spaces, the following two conditions are equivalent:*

1. The product space $\prod_{i \in I} X_i$ is a \mathbf{K} -space.
2. For each $i \in I$, X_i is a \mathbf{K} -space.

In particular, we have the following.

Corollary 7.13. ([29, 42, 44]) *For any family $\{X_i : i \in I\}$ of T_0 spaces, the following two conditions are equivalent:*

1. *The product space $\prod_{i \in I} X_i$ is well-filtered.*
2. *For each $i \in I$, X_i is well-filtered.*

8. \mathbf{DCPO}_s -completions of posets

Let **Poset** denote the category of all posets with monotone (i.e. order-preserving) mappings, \mathbf{DCPO}_s the category of all dcpos with Scott continuous mappings, and \mathbf{Poset}_s the category of all posets with Scott continuous mappings. Then \mathbf{DCPO}_s is a full subcategory of \mathbf{Poset}_s .

Definition 8.1. ([50]) A \mathbf{DCPO}_s -completion of a poset P , \mathbf{D}_s -completion of P for short, is a pair $\langle \tilde{P}, \eta \rangle$ consisting of a dcpo \tilde{P} and a Scott continuous mapping $\eta : P \rightarrow \tilde{P}$, such that for any Scott continuous mapping $f : P \rightarrow Q$ to a dcpo Q , there exists a unique Scott continuous mapping $\tilde{f} : \tilde{P} \rightarrow Q$ such that $\tilde{f} \circ \eta = f$, that is, the following diagram commutes.

\mathbf{D}_s -completions, if they exist, are unique up to isomorphism. We shall use $\mathbf{D}_s(P)$ to denote the \mathbf{D}_s -completion of P if it exists.

In [50], using the D -topologies and d -closure defined in [50] (see also [19]), which originates from Wyler [32], Zhao and Fan proved that for any poset P , the \mathbf{D}_s -completion of P exists. As Keimel and Lawson pointed out in [19] that the \mathbf{D}_s -completion of a poset P is a special case of the d -reflection of a certain T_0 space. More precisely, the d -reflection of Scott space ΣP .

Proposition 8.2. ([36]) *For a poset P , $\mathbf{D}_s(P) = d(\Sigma P)$ with the canonical mapping $\eta_P : P \rightarrow \mathbf{D}_s(P)$, $\eta_P(x) = \text{cl}_{\sigma(P)}\{x\}$, is the \mathbf{D}_s -completion of P .*

Corollary 8.3. ([36, 50]) \mathbf{DCPO}_s is reflective in \mathbf{Poset}_s .

Remark 8.4. ([39]) The \mathbf{DCPO}_s -completion (that is, \mathbf{DCPO}_s -completion) was called the \mathbf{D} -completion in ([50], Definition 1) and the \mathbf{DCPO} -completion in [36]. For the sake of distinction, here we call such a completion the \mathbf{DCPO}_s -completion and give the \mathbf{DCPO} -completion (or \mathbf{D} -completion) a little different meaning (see Section 9).

By Propositions 7.5 and 8.2, we get the following corollary, which generalizes ([19], Proposition 7.8).

Proposition 8.5. ([38]) *For a family $\{P_i : i \in I\}$ of posets, the product $\prod_{i \in I} \mathbf{D}_s(P_i)$ is the \mathbf{D}_s -completion of the product $\prod_{i \in I} P_i$, provided the Scott topology of the product $\prod_{i \in I} P_i$ is the product of the Scott topologies of the factors.*

9. Scott \mathbf{K} -completions of posets

Definition 9.1. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing **Sob**. A poset P is called a *Scott \mathbf{K} -dcpo*, a \mathbf{K} -dcpo for short, if ΣP is a \mathbf{K} -space. A poset (even a dcpo) Q is said to be a *non- \mathbf{K} dcpo* if Q is not a \mathbf{K} -dcpo. Let $\mathbf{K-DCPO}_s$ denote the category of all \mathbf{K} -dcpos with Scott continuous mappings.

$\mathbf{K}\text{-DCPO}_s$ is a full subcategory of \mathbf{DCPO}_s , and it is a subcategory of \mathbf{DCPO} , but not a full subcategory of \mathbf{DCPO} .

Clearly, a poset P is a \mathbf{Top}_d -dcpo (\mathbf{d} -dcpo for short) iff P is a dcpo. For $\mathbf{K} = \mathbf{Top}_w$, the \mathbf{K} -dcpos are simply called the \mathbf{WF} -dcpos, and the category $\mathbf{Top}_w\text{-DCPO}_s$ is simply denoted as $\mathbf{WF}\text{-DCPO}_s$.

Definition 9.2. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing \mathbf{Sob} . A Scott \mathbf{K} -completion, \mathbf{K}_s -completion for short, of a poset P is a pair $\langle \tilde{P}, \eta \rangle$ consisting of a \mathbf{K} -dcpo \tilde{P} and a Scott continuous mapping $\eta : P \rightarrow \tilde{P}$, such that for any Scott continuous mapping $f : P \rightarrow Q$ to a \mathbf{K} -dcpo Q , there exists a unique Scott continuous mapping $\tilde{f} : \tilde{P} \rightarrow Q$ such that $\tilde{f} \circ \eta = f$, that is, the following diagram commutes.

When $\mathbf{K} = \mathbf{Top}_d$, the \mathbf{K}_s -completion is the \mathbf{DCPO}_s -completion. For $\mathbf{K} = \mathbf{Top}_w$, the \mathbf{K}_s -completion is simply called the \mathbf{WF}_s -completion.

By a standard argument, \mathbf{K}_s -completions, if they exist, are unique up to isomorphism. We use $\mathbf{K}_s(P)$ to denote the \mathbf{K}_s -completion of P if it exists. We will use $\mathbf{D}_s(P)$, $\mathbf{WF}_s(P)$ and $\mathbf{Sob}_s(P)$ to denote the \mathbf{DCPO}_s -completion, \mathbf{WF}_s -completion, and \mathbf{Sob}_s -completion of P , respectively.

Definition 9.3. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing \mathbf{Sob} . A \mathbf{K} -completion of a poset P is a pair $\langle \tilde{P}, \phi \rangle$ consisting of a \mathbf{K} -dcpo \tilde{P} and a monotone mapping $\phi : P \rightarrow \tilde{P}$, such that for any monotone mapping $f : P \rightarrow Q$ to a \mathbf{K} -dcpo Q , there exists a unique Scott continuous mapping $\tilde{f} : \tilde{P} \rightarrow Q$ such that $\tilde{f} \circ \phi = f$.

When $\mathbf{K} = \mathbf{Top}_d$ the \mathbf{K} -completion is the \mathbf{DCPO} -completion. For $\mathbf{K} = \mathbf{Top}_w$, the \mathbf{K} -completion is simply called the \mathbf{WF} -completion.

Similarly, \mathbf{K} -completions, if they exist, are unique up to isomorphism. We use $\mathbf{K}(P)$ to denote the \mathbf{K} -completion of P if it exists. We will use $\mathbf{D}(P)$, $\mathbf{WF}(P)$ and $\mathbf{Sob}(P)$ to denote the \mathbf{DCPO} -completion of P , \mathbf{WF} -completion, and \mathbf{Sob} -completion of P , respectively.

Definition 9.4. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing \mathbf{Sob} . A poset P is called a weak \mathbf{K} -dcpo if there is a \mathbf{K} -space such that P is isomorphic to ΩX .

Clearly, every weak \mathbf{K} -dcpo is a dcpo, and a poset P is a dcpo iff P is a \mathbf{d} -dcpo iff P is a weak \mathbf{d} -dcpo. By Corollary 3.3, every complete lattice is a \mathbf{WF} -dcpo and is also a weak \mathbf{Sob} -dcpo. The Isbell's lattice L , as a complete lattice, is a weak \mathbf{Sob} -dcpo but not a \mathbf{Sob} -dcpo.

Theorem 9.5. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing \mathbf{Sob} which is adequate and closed with respect to homeomorphisms. For a poset P , if $\mathbf{K}(\Sigma P)$ is a \mathbf{K} -dcpo, then $\mathbf{K}_s(P) = \mathbf{K}(\Sigma P)$ with the canonical mapping $\eta_P : P \rightarrow \mathbf{K}_s(P)$, $\eta_P(x) = \text{cl}_{\sigma(P)}\{x\} = \downarrow x$, is a \mathbf{K}_s -completion of P .

Definition 9.6. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_0 . A poset P is called a $\mathbf{S_K}$ -poset if $\Sigma \mathbf{K}(\Sigma P)$ is a \mathbf{K} -space. Let $\mathbf{S_K}\text{-Poset}_s$ denote the category of all $\mathbf{S_K}$ -posets with Scott continuous mappings.

Clearly, $\mathbf{S_K}\text{-Poset}_s$ is a full subcategory of \mathbf{Poset}_s . If \mathbf{K} is a full subcategory of \mathbf{Top}_d containing \mathbf{Sob} , then by Proposition 4.7, every \mathbf{K} -dcpo is a $\mathbf{S_K}$ -poset, and hence $\mathbf{K}\text{-DCPO}_s$ is a full subcategory of $\mathbf{S_K}\text{-Poset}_s$.

From Theorem 9.5 we deduce the following result.

Corollary 9.7. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing \mathbf{Sob} which is adequate and closed with respect to homeomorphisms. Then $\mathbf{K}\text{-DCPO}_s$ is reflective in $\mathbf{S_K}\text{-Poset}_s$. Therefore, if $\mathbf{K}(\Sigma P)$ is a \mathbf{K} -dcpo for any poset P , then $\mathbf{K}\text{-DCPO}_s$ is reflective in \mathbf{Poset}_s .

By Theorems 5.6 and 9.5, we get the following three corollaries.

Corollary 9.8. ([39]) For a poset P , if $\text{Irr}_c(\Sigma P)$ is a **Sob**-dcpo, then $\mathbf{Sob}_s(P) = \text{Irr}_c(\Sigma P)$ with the canonical mapping $\eta_P : P \rightarrow \mathbf{Sob}_s(P)$, $\eta_P(x) = \text{cl}_c \sigma(P)\{x\} = \downarrow x$, is a **Sob**_s-completion of P .

Corollary 9.9. ([39]) For a poset P , if $\text{WD}(\Sigma P)$ is a **WF**-dcpo, then $\mathbf{WF}_s(P) = \text{WD}(\Sigma P)$ with the canonical mapping $\eta_P : P \rightarrow \mathbf{WF}_s(P)$, $\eta_P(x) = \text{cl}_{\sigma(P)}\{x\} = \downarrow x$, is a **WF**_s-completion of P .

Corollary 9.10. Let \mathbf{K} be a Keimel-Lawson category which is a full subcategory of \mathbf{Top}_d . For a poset P , if $\mathbf{K}(\Sigma P)$ is a **K**-dcpo, then $\mathbf{K}_s(P) = \mathbf{K}(\Sigma P)$ with the canonical mapping $\eta_P : P \rightarrow \mathbf{K}_s(P)$, $\eta_P(x) = \text{cl}_{\sigma(P)}\{x\} = \downarrow x$, is a **K**_s-completion of P .

Proposition 9.11. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing **Sob** which is adequate and closed with respect to homeomorphisms. For a non-**K** poset P , if $\text{Irr}_c(\Sigma P) = \{\overline{\{x\}} : x \in P\} \cup \{P\}$ and P_T is a **K**-dcpo, then $\mathbf{K}_s(P) = P_T$ with the canonical mapping $\eta_P : P \rightarrow P_T$, $\eta_P(x) = x$, is a **K**_s-completion of P .

By Corollary 6.9 and Theorem 9.5, we get the following result.

Corollary 9.12. ([39]) Let P be a poset P . If $\text{Irr}_c(\Sigma P) = \{\overline{\{x\}} : x \in P\} \cup \{P\}$ and ΣP_T is a sober space, then $\mathbf{Sob}_s(P) = P_T$ with the canonical mapping $\eta_P : P \rightarrow P_T$, $\eta_P(x) = x$, is a **Sob**_s-completion of P .

Corollary 9.13. ([39]) Let P be a poset P . If $\text{Irr}_c(\Sigma P) = \{\overline{\{x\}} : x \in P\} \cup \{P\}$ and ΣP_T is a well-filtered space, then $\mathbf{WF}_s(P) = P_T$ with the canonical mapping $\eta_P : P \rightarrow P_T$, $\eta_P(x) = x$, is a **WF**_s-completion of P .

10. K-reflections of Alexandroff spaces

In this section, we discuss the **K**-reflections of Alexandroff spaces and the **K**-completions of posets.

It is straightforward to verify the following lemma.

Lemma 10.1. ([39]) For a poset P , a T_0 space Y , and a mapping $f : (P, \alpha(P)) \rightarrow Y$, the following conditions are equivalent:

1. $f : (P, \alpha(P)) \rightarrow Y$ is continuous.
2. $f : P \rightarrow \Omega Y$ is monotone.
3. $f : (P, \alpha(P)) \rightarrow (\Omega Y, \alpha(\Omega Y))$ is continuous.

Lemma 10.2. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing **Sob** and P a poset. Then $\mathbf{K}((P, \alpha(P))) = \text{Id } P$.

Theorem 10.3. ([39]) Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing **Sob** which is adequate and closed with respect to homeomorphisms. Then for any poset P , the **K**-reflection $(\)^{\mathbf{K}}$ of $(P, \alpha(P))$ exists and it is a Scott space. More precisely, the Scott space $\Sigma \text{Id } P$ with the canonical mapping $\phi_P : (P, \alpha(P)) \rightarrow \Sigma \text{Id } P$, $x \mapsto \text{cl}_{\gamma(P)}\{x\} = \downarrow x$, is a **K**-reflection of $(P, \alpha(P))$.

Remark 10.4. ([39]) We can present a direct proof of Theorem 10.3.

Proof. Since $\text{Id } P$ is an algebraic domain (cf. [4], Proposition I-4.10), by ([4], Proposition III-3.7) and Lemma 10.2, $\mathbf{K}((P, \alpha(P))) = \text{Id } P$ and $\Sigma \text{Id } P$ is sober and hence

a \mathbf{K} -space. Clearly, the canonical mapping $\phi_P : (P, \alpha(P)) \rightarrow \Sigma \mathbf{K}((P, \alpha(P))) = \Sigma \text{Id } P$, $x \mapsto \text{cl}_{\alpha(P)}\{x\} = \downarrow x$, is continuous. Now we show that for each \mathbf{K} -space Y and each continuous mapping $f : (P, \alpha(P)) \rightarrow Y$, there is a unique continuous mapping $f^* : \Sigma \text{Id } P \rightarrow Y$ such that $f^* \circ \phi_P = f$, that is, the following diagram commutes.

We first prove the existence of f^* . Since Y is a \mathbf{K} -space, Y is a d -space. Therefore, $\vee E$ exists in Y for each directed subset E of Y (with the specialization order). As $f : (P, \alpha(P)) \rightarrow Y$ is continuous, $f : P \rightarrow \Omega Y$ is monotone. Define a mapping $f^* : \Sigma \text{Id } P \rightarrow Y$ by $f^*(I) = \vee f(I)$ for each $I \in \text{Id } P$. For every $\{I_d : d \in D\} \in \mathcal{D}(\text{Id } P)$, we have that $f^*(\vee_{\text{Id } P} \{I_d : d \in D\}) = f^*(\cup_{d \in D} I_d) = \vee f(\cup_{d \in D} I_d) = \vee_{d \in D} \vee f(I_d) = \vee_{d \in D} f^*(I_d)$. By Lemmas 2.2 and 2.4, $f^* : \Sigma \text{Id } P \rightarrow Y$ is continuous. For each $x \in P$, since $f : P \rightarrow \Omega Y$ is monotone, we have $f^*(\phi_P(x)) = \vee f(\downarrow x) = f(x)$, proving that $f^* \circ \phi_P = f$.

Now we prove the uniqueness of f^* . Suppose that $g : (P, \alpha(P)) \rightarrow Y$ is another continuous mapping satisfying $g \circ \phi_P = f$. Then for each $I \in \text{Id } P$, by Lemmas 2.2 and 2.4, we have that

$$g(I) = g(\cup_{x \in I} \downarrow x) = g(\vee_{\text{Id } P} \{\downarrow x : x \in I\}) = \vee_{x \in I} g(\downarrow x) = \vee_{x \in I} f(x) = \vee f(I) = f^*(I),$$

and hence $g = f^*$.

Therefore, $(\Sigma \text{Id } P, \phi_P : (P, \alpha(P)) \rightarrow \Sigma \text{Id } P)$ is a \mathbf{K} -reflection of $(P, \alpha(P))$. \square

From Theorem 10.3, we immediately deduce the following result.

Corollary 10.5. ([39]) *Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing **Sob** which is adequate and closed with respect to homeomorphisms. Then for any poset P , the \mathbf{K} -completion of P exists and it is the pair $\langle \text{Id } P, \phi_P \rangle$, where $\phi_P : P \rightarrow \text{Id } P$ is defined by $\phi_P(x) = \downarrow x$ for each $x \in P$.*

Corollary 10.6. ([39]) *Let \mathbf{K} be a full subcategory of \mathbf{Top}_d containing **Sob** which is adequate and closed with respect to homeomorphisms. Then $\mathbf{K}\text{-DCPO}_s$ is reflective in **Poset**.*

Finally, by Theorem 5.6 and Corollaries 10.5 and 10.6, we have the following three corollaries.

Corollary 10.7. ([39]) *Let P be a poset. Then the d -reflection of ΓP , the well-filtered reflection of ΓP and the sobrification of ΓP agree. They all are the Scott space ΣId with the canonical mapping $\phi_P : \Gamma P \rightarrow P$.*

Corollary 10.8. ([39]) *Let P be a poset. Then the \mathbf{D} -completion of P , the **WF**-completion of P and the **Sob**-completion of P agree. They all are the pair $\langle \text{Id } P, \phi_P \rangle$, where $\phi_P : P \rightarrow \text{Id } P$ is defined by $\phi_P(x) = \downarrow x$ for each $x \in P$.*

Corollary 10.9. ([39]) *DCPO_s , $\mathbf{WF}\text{-DCPO}_s$ and $\mathbf{Sob}\text{-DCPO}_s$ all are reflective in **Poset**.*

11. Sobriety of Scott topology on dcpos

The following concept was first introduced in ([35], Definition 10.2.11) (see also [30]).

Definition 11.1. ([35]) A T_0 space X is said to have *property R* if for any family $\{\uparrow F_i : i \in I\} \subseteq \mathbf{Fin } P$ and any $U \in \mathcal{O}(X)$, $\cap_{i \in I} \uparrow F_i \subseteq U$ implies $\cap_{i \in I_0} \uparrow F_i \subseteq U$ for some $I_0 \in I^{(<\omega)}$. For a poset P , when ΣP has property R, we will simply say that P has property R.

Proposition 11.2. ([35, 40]) *For a poset P , consider the following conditions:*

1. $(P, \lambda(P))$ is compact.
2. $(P, \lambda(P))$ is upper semicompact.

3. P has property R .

4. For any filtered family $\{\uparrow F_i : i \in I\} \subseteq \mathbf{Fin} P$ and any $U \in \sigma(P)$, $\cap_{i \in I} \uparrow F_i \subseteq U$ implies $\uparrow F_j \subseteq U$ for some $j \in I$.

5. P is a dcpo.

Then $(1) \Rightarrow (2) \Rightarrow (3) \Rightarrow (4) \Rightarrow (5)$.

Theorem 11.3. ([35, 40]) For a poset P , the following conditions are equivalent:

1. P has property R .

2. For any $\{x_i : i \in I\} \subseteq P$ and any $U \in \sigma(P)$ with $\cap_{i \in I} \uparrow x_i \subseteq U$, there is $I_0 \in I^{(<\omega)}$ such that $\cap_{i \in I_0} \uparrow x_i \subseteq U$.

3. The compact saturated subsets of $(P, \omega(P))$ are exactly the closed subsets of ΣP .

4. Every Scott closet subset of P is compact in ΣP .

Proposition 11.4. ([35, 40]) Let P be a poset. If ΣP is compact and $\cap_{x \in F} \uparrow x$ is compact in ΣP for each $F \in P^{(<\omega)}$, then the following two conditions are equivalent:

1. $(P, \lambda(P))$ is compact.

2. P has property R .

Clearly, if $(P, \lambda(P))$ is compact, then $(P, \sigma(P))$ is compact, and for each $F = \{x_1, x_2, \dots, x_n\} \in P^{(<\omega)}$, $\uparrow x_1 \cap \uparrow x_2 \cap \dots \cap \uparrow x_n$ is a closed subset of $(P, \lambda(P))$, whence it is compact in $(P, \lambda(P))$, and consequently, compact in $(P, \sigma(P))$. So Proposition 11.4 can be restated as the following one.

Proposition 11.5. ([35, 40]) For a poset, the following two conditions are equivalent:

1. $(P, \lambda(P))$ is compact.

2. ΣP is compact, $\uparrow x_1 \cap \uparrow x_2 \cap \dots \cap \uparrow x_n$ is Scott compact (i.e., compact in ΣP) for each $\{x_1, x_2, \dots, x_n\} \in P^{(<\omega)}$, and P has property R .

The following concept was introduced by Lawson, Wu, and Xi [21].

Definition 11.6. ([21]) A T_0 space X is said to be Ω^* -compact if every closed subset of X is compact in $(X, \omega(X))$.

As $\{X \setminus \uparrow F : F \in \mathbf{Fin} X\}$ is a base of $\omega(X)$, we clearly have the following result.

Proposition 11.7. Let X be a T_0 space. Then the following two conditions are equivalent:

1. X has property R .

2. X is Ω^* -compact.

Proposition 11.8. ([40]) Let P be a dcpo for which ΣP is well-filtered and coherent. Then P has property R .

In what follows, for a poset P , let $\omega^*(P)$ denote the family of all closed subsets of $(P, \omega(P))$, namely, $\omega^*(P) = \{\emptyset, P\} \cup \{\cap \mathcal{F} : \mathcal{F} \subseteq \mathbf{Fin}P\}$.

Lemma 11.9. ([40]) *Let P be a poset and $Q = (\omega^*(P), \supseteq)$ (i.e., the order on Q is the reverse inclusion order). Then $m : P \times P \rightarrow Q$, $(x, y) \mapsto \uparrow x \cap \uparrow y$, is Scott continuous.*

Corollary 11.10. ([40]) *Let P be a poset and $Q = (\omega^*(P), \supseteq)$. If $\Sigma(P \times P) = \Sigma P \times \Sigma P$, then $m : \Sigma P \times \Sigma P \rightarrow \Sigma Q$, $m(x, y) = \uparrow x \cap \uparrow y$, is continuous.*

Corollary 11.11. ([40]) *Let P be a poset and $Q = (\omega^*(P), \supseteq)$. If ΣP is core compact (especially, locally compact), then $m : \Sigma P \times \Sigma P \rightarrow \Sigma Q$, $m(x, y) = \uparrow x \cap \uparrow y$, is continuous.*

Using the Scott topology and the lower topology, we give the following useful characterization of property R.

Lemma 11.12. ([40]) *Let P be a poset and $Q = (\omega^*(P), \supseteq)$. Then the following two conditions are equivalent:*

1. P has property R.
2. For any $U \in \sigma(P)$, $\Phi(U) = \{C \in \omega^*(P) : C \subseteq U\} \in \sigma(Q)$.

Theorem 11.13. ([40]) *Let P be a poset satisfying property R and $\Sigma(P \times P) = \Sigma P \times \Sigma P$. Then ΣP is sober.*

Corollary 11.14. ([40]) *Let P be a poset satisfying $\Sigma(P \times P) = \Sigma P \times \Sigma P$. If ΣP is well-filtered and coherent. Then ΣP is sober.*

Clasification


2000 MSC: 54D99; 54B30; 54B20; 06B35

Author details

Xiaoquan Xu
School of Mathematics and Statistics, Minnan Normal University, Zhangzhou, China

*Address all correspondence to: xiqxu2002@163.com

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Abramsky S, Jung A. Domain theory, semantic structures. In: Abramsky S, Gabbay D, Maibaum T, editors. *Handbook of Logic in Computer Science*. Vol. 3. Oxford: Clarendon Press; 1994. pp. 1-168
- [2] Ershov Y. On d -spaces. *Theoretical Computer Science*. 1999;**224**:59-72
- [3] Ershov Y. The d -rank of a topological space. *Algebra and Logic*. 2017;**56**:98-107
- [4] Gierz G, Hofmann K, Keimel K, Lawson J, Mislove M, Scott D. Continuous Lattices and Domains. In: *Encycl. Math. Appl.* Vol. 93. Cambridge University Press; 2003
- [5] Gierz G, Lawson J. Generalized continuous and hypercontinuous lattices. *Rocky Mountain Journal of Mathematics*. 1981;**11**:271-296
- [6] Gierz G, Lawson J, Stralka A. Quasicontinuous posets. *Houston Journal of Mathematics*. 1983;**9**:191-208
- [7] Goubault-Larrecq J. *Non-Hausdorff Topology and Domain Theory*, New Mathematical Monographs. Vol. 22. New York: Cambridge University Press; 2013
- [8] Heckmann R. An upper power domain construction in terms of strongly compact sets, *Lecture Notes in Computer Science*. Vol. 598. Berlin Heidelberg, New York: Springer; 1992. pp. 272-293
- [9] Heckmann R, Keimel K. Quasicontinuous domains and the Smyth powerdomain. *Electronic Notes in Theoretical Computer Science*. 2013;**298**: 215-232
- [10] Hertling P. Topological properties of the binary supremum function. *Semigroup Forum*. 2022;**104**:618-646
- [11] Hoffmann R. On the sobrification remainder ${}^sX - X$. *Pacific Journal of Mathematics*. 1979;**83**:145-156
- [12] Hoffmann R. Sobrification of partially ordered sets. *Semigroup Forum*. 1979;**17**:123-138
- [13] Hofmann K, Lawson J. The spectral theory of distributive continuous lattices. *Transactions of the American Mathematical Society*. 1978;**246**:285-310
- [14] Hofmann K, Mislove M. Local compactness and continuous lattices. *Lecture Notes in Mathematics*. 1981;**871**: 125-158
- [15] Isbell J. Completion of a construction of Johnstone. *Proceedings of the American Mathematical Society*. 1982;**85**:333-334
- [16] Jia X. Meet-Continuity and Locally Compact Sober Dcpo's [thesis]. Birmingham, England: University of Birmingham; 2018
- [17] Johnstone P. Scott is not always sober. In: *Continuous Lattices*, *Lecture Notes in Math.* Vol. 871. Berlin: Springer-Verlag; 1981. pp. 282-283
- [18] Jung A. Four dcpo's, a theorem, and an open problem, preprint
- [19] Keimel K, Lawson J. D -completions and the d -topology. *Annals of Pure and Applied Logic*. 2009;**159**(3):292-306
- [20] Kou H. U_k -admitting dcpo's need not be sober. In: *Domains and Processes, Semantic Structure on Domain Theory*. Vol. 1. Dordrecht: Kluwer Academic Publishers; 2001. pp. 41-50
- [21] Lawson J, Wu G, Xi X. Well-filtered spaces, compactness, and the lower

topology. Houston Journal of Mathematics. 2020;**46**(1):283-294

[22] Liu B, Li Q, Wu G. Well-filterifications of topological spaces. Topology and Its Applications. 2020;**279**:107245

[23] Lu J, Wang K, Wu G, Zhao B. Nonexistence of k -bounded sobrifcation to appear. Houston Journal of Mathematics. 2022

[24] Lyu Z, Chen Y, Jia X. Core-compactness, consonance and the Smyth powerspaces. Topology and Its Applications. 2022;**312**:108066

[25] Miao H, Li Q, Zhao D. On two problems about sobriety of topological spaces. Topology and Its Applications. 2021;**295**:107667

[26] Miao H, Xi X, Li Q, Zhao D. Not every countable complete lattice is sober. arXiv:2205.00250

[27] Schalk A. Algebras for Generalized Power Constructions [thesis]. Darmstadt: Technische Hochschule Darmstadt; 1993

[28] Shen C, Wu G, Xi X, Zhao D. Sober Scott spaces are not always co-sober. Topology and Its Applications. 2019;**282**:107316

[29] Shen C, Xi X, Xu X, Zhao D. On well-filtered reflections of T_0 spaces. Topology and Its Applications. 2019;**267**:106869

[30] Wen X, Xu X. Sober is not always co-sober. Topology and Its Applications. 2018;**250**:48-52

[31] Wu G, Xi X, Xu X, Zhao D. Existence of well-filterification. Topology and Its Applications. 2019;**267**:107044

[32] Wyler U. Dedekind complete posets and Scott topologies. Lecture Notes in Mathematics. 1981;**871**:384-389

[33] Xi X, Lawson J. On well-filtered spaces and ordered sets. Topology and Its Applications. 2017;**228**:139-144

[34] Xi X, Zhao D. Well-filtered spaces and their dcpo models. Mathematical Structures in Computer Science. 2017;**27**:507-515

[35] Xu X. Order and Topology. Beijing: Science Press; 2016

[36] Xu X. A direct approach to K-reflections of T_0 spaces. Topology and Its Applications. 2020;**272**:107076

[37] Xu X. On H-sober spaces and H-sobifications of T_0 spaces. Topology and Its Applications. 2021;**289**:107548

[38] Xu X. K-reflections of product spaces. Topology and Its Applications. 2021;**289**:107571

[39] Xu X. K-reflections of Scott space. arXiv:2204.09512. 2022

[40] Xu X. On sobriety of Scott topology on dcpos. arXiv:2211.15027. 2022

[41] Xu X, Shen C, Xi X, Zhao D. First countability, ω -well-filtered spaces and reflections. Topology and Its Applications. 2020;**279**:107255

[42] Xu X, Shen C, Xi X, Zhao D. On T_0 spaces determined by well-filtered spaces. Topology and Its Applications. 2020;**282**:107323

[43] Xu X, Shen C, Xi X, Zhao D. First-countability, ω -Rudin spaces and well-filtered determined spaces. Topology and Its Applications. 2021;**300**:107775

- [44] Xu X, Xi X, Zhao D. A complete Heyting algebra whose Scott topology is not sober. *Fundamenta Mathematicae*. 2021;**352**:315-323
- [45] Xu X, Yang Z. Coincidence of the upper Vietoris topology and the Scott topology. *Topology and Its Applications*. 2021;**288**:107480
- [46] Xu X, Zhao D. On topological Rudin's lemma, well-filtered spaces and sober spaces. *Topology and Its Applications*. 2020;**272**:107080
- [47] Xu X, Zhao D. Some open problems on well-filtered spaces and sober spaces. *Topology and Its Applications*. 2021;**301**: 107540
- [48] Zhang Z, Li Q. A direct characterization of the monotone convergence space completion. *Topology and Its Applications*. 2017;**230**: 99-104
- [49] Zhao B, Lu J, Wang K. The answer to a problem posed by Zhao and Ho. *Acta Mathematica Sinica*. 2019;**35**(3):438-444
- [50] Zhao D, Fan T. Dcpo-completion of posets. *Theoretical Computer Science*. 2010;**411**:2167-2173
- [51] Zhao D, Ho W. On topologies defined by irreducible sets. *Journal of Logical and Algebraic Methods in Programming*. 2015;**84**(1):185-195
- [52] Zhao D, Xi X, Chen Y. A new dcpo whose Scott topology is well-filtered but not sober. *Topology and Its Applications*. 2019;**252**:97-02
- [53] Engelking R. *General Topology*. Warszawa: Polish Scientific Publishers; 1989
- [54] Ern  M. *Categories of locally hypercompact spaces and quasicontinuous posets*. *Applied Categorical Structures*. 2018;**26**:823-854
- [55] Ern  M. Infinite distributive laws versus local connectedness and compactness properties. *Topology and Its Applications*. 2009;**156**:2054-2069
- [56] Ern  M, Gatzke H. Convergence and continuity in partially ordered sets and semilattices. In: Hoffmann R, Hofmann K, editors. *Continuous Lattices and Their Applications*, *Lecture Notes in Pure and Appl. Math.* Vol. 101. New York: Marcel Dekker; 1985. pp. 9-40
- [57] Levy A. *Basic Set Theory*. Mineola, New York: Dover Publications, Inc.; 2002

The Topological Connectivity of the Bipartite Graph's Independence Complex

Yousef Methkal Abd Algani and Amal Sharif-Rasslan

Abstract

Let (\mathcal{C}) signify a simplicial complex; the topological connectivity of \mathcal{C} is denoted by $\zeta(\mathcal{C})$, which equals $\mathcal{C}+2$, where ζ is a function from the class of graphs to the set $\mathbb{Z}^+ = \{0, 1, 2, \dots, \infty\}$. Let the function ζ possesses the following characteristics: 1. $\zeta(K_0) = 0$. 2. For any graph $G(V, E)$, there exists an edge $\{x, y\}$ in G such that $\zeta(G(V, E) - \{x, y\}) \geq \zeta(G(V, E))$ where $(G(V, E) - \{x, y\})$ means to remove the edge $\{x, y\}$ from the graph $G(V, E)$, and $\zeta(G(V, E) - \Delta(\{x, y\})) \geq \zeta(G(V, E)) - 1$ where $(G(V, E) - \Delta(\{x, y\}))$ means to remove the vertices x and y and all their neighbors. Then $\mu(\Gamma(G(V, E))) \geq \zeta(G(V, E))$. Let ζ_0 denote the maximal function ζ that satisfies the aforementioned constraints. Berger proved that $\mu(\Gamma(G(V, E))) \geq \zeta_0(G(V, E))$ for trees and for complements of chordal graphs. Kawamura proved the same theorem for chordal graphs, and Abd Algani proved it for circular-arc graph: Let $G(V, E)$ be a circular-arc graph; if $\zeta_0(G(V, E)) \leq 2$, then $\mu(\Gamma(G(V, E))) \leq 2$. In this chapter, we will prove the following theorem: Let $G(V, E)$ be a bipartite graph; if $\zeta_0(G(V, E)) \leq 2$, then $\mu(\Gamma(G(V, E))) \leq 2$.

Keywords: topological connectivity, independence complex, bipartite graph, bipartite graph's independence complex, chordal graph

1. Introduction

The set $\mathcal{C} \neq \emptyset$ is a simplicial complex if $\omega \in \mathcal{C}$ and $\lambda \subseteq \omega$ indicate $\lambda \in \mathcal{C}$. Each simplicial complex has a distinctive (up to homeomorphism) geometric recognition, specifically an inserting in the space \mathbb{R}^n , wherein each simplex $\omega \in \mathcal{C}$ is grasped as a homeomorph of a simplex in \mathbb{R}^n .

The connection between two simplicial complexes M and N is called join, and it will be denoted by $M * N$ and is defined as

$$M * N = \{\omega \cup \lambda \mid \omega \in M, \lambda \in N\}.$$

The topological space $X * v_{(\text{vertex})}$ is called a cone over X and is denoted by $\mathcal{C}_v X$. And the topological space $X * S_0$ is called a suspension and is denoted by $\text{susp} - M$.

A complex would be identified by its geometric recognition. The topological connectivity of a simplicial complex \mathbb{C} is the biggest number \mathbb{C}_0 that being said for each number $d \leq \mathbb{C}_0$, each inserting of the d -dimensional sphere S_d is extendable to an inserting of the $(d + 1)$ -dimensional ball B_{d+1} in \mathbb{C} . The connectivity of a complex could be infinite. The connectivity of $\mathbb{C} + 2$ is signified by $\mu(\mathbb{C})$. This is because the addition of 2 makes the creation of some grades more sophisticated.

For a graph $G(V, E)$ with the vertex set V and edge set E , a subgroup U of V is no two distinct vertices of U in V are adjacent. A subgroup S of V is leading in $G(V, E)$ if each vertex $v \in V$ is nearby to a vertex of S . The dominance number $\delta(G(V, E))$ is the lowest of the cardinality of leading groups of

$G(V, E) : \delta(G(V, E)) = \min\{|S| : S \text{ is dominating in } G(V, E)\}.$

Let H be a graph; we signify by $\Gamma(H)$ the simplicial complex containing all the autonomous groups of vertices in H . Let $V_k, k \leq m$, be a divider of the vertex set of a graph G . Given a vertex $v \in V(G(V, E))$, we write $k(v)$ for the index k for which $v \in V_k$. For a set Z of vertices, we write $\Gamma(Z) = \{k(z) : z \in Z\}$. Typically, we signify the set $\{1, \dots, f\}$ by $[f]$. Assuming a subset Γ of $[f]$, we write V_Γ for $\bigcup_{k \in \Gamma} V_k$.

Consider a graph $G(V, E)$ with a divider V_1, V_2, \dots, V_f of its vertex set. A select of one vertex from every set V_i is termed an autonomous structure of representative (ISR) if the designated vertices are nonadjacent in $G(V, E)$. The connection between topological connectivity and ISR was recognized in [1]. Existence of ISR:

Theorem 1. If for all $\Gamma \subseteq [f]$

$$\mu(\Gamma(G(V, E)[V_\Gamma])) \geq |\Gamma|,$$

then the partition $(V_k) (k \leq f)$ of $V(G(V, E))$ has an ISR.

Graph $G(V, E)$ is a bipartite graph if there is a division of V into two groups, A and B , so that every edge of $G(V, E)$ has one end at A and the other at B . In this instance, there occurs an ISR if and only if the graph is not a whole bipartite. But not being complete bipartite means the presence of a linking in $\Gamma(G(V, E))$ between both simplices V_1 and V_2 of $\Gamma(G(V, E))$. Consequently, not being a whole bipartite is equal to $\Gamma(G(V, E))$ being related, which, in the aforementioned terminology, means being *zero-connected*, which means that $\mu(\Gamma(G(V, E))) \geq 2$. Consequently, in this instance, the state of the theorem is not only adequate but also essential. The following theorem is due to Ahrone [1]; we carry here its new formulation.

Theorem 2. Let (\mathbb{C}) signify a simplicial complex and $\zeta(\mathbb{C})$ denote the topological connectivity of \mathbb{C} , which equals $\mathbb{C} + 2$. Let ζ be a function from the class of graphs to the set $\mathbb{Z}^+ = \{0, 1, 2, \dots, \infty\}$.

Let the function ζ possess the following characteristics:

$$1. \zeta(K_0) = 0$$

2. There exists, for any graph G , an edge $\{x, y\}$ in G ,

$$\zeta(G(V, E) - \{x, y\}) \geq \zeta(G(V, E))$$

$(G(V, E) - \{x, y\})$ means to remove the edge $\{x, y\}$ from the graph $G(V, E)$,
and

$$\zeta(G(V, E) - \Delta(\{x, y\})) \geq \zeta(G(V, E)) - 1.$$

Then

$$\mu(\Gamma(G(V, E))) \geq \zeta(G(V, E)).$$

$(G(V, E) - \Delta(\{x, y\}))$ means to remove the vertices x and y and all their neighbors).

Let ζ_0 denote the maximal function ζ that satisfies the aforementioned constraints.

Actually, there exists a maximal function ζ_0 filling the conditions of the theorem. It is best defined in terms of a game between two actors: (a) and (b). Player (a) needs to take full benefit of the function ζ in the theorem (and hereafter show that ζ is big), whereas player (b) needs to minimize ζ . Player (a) chooses an edge $e = \{x, y\}$ in the graph assumed at the current phase of the game. Player (b) selects between two options: he or she either (I) removes e from the graph, or else (II) erases all neighbors of the vertices x and y (including, x and y themselves). The game finishes when either there remains a secluded vertex, in which case ζ is well-defined as ∞ , or there are no remaining vertices, in which case ζ is recognized as the number of steps of player (b) of type (II).

We describe $\zeta_0(G(V, E))$ as the maximal value of $\zeta(G(V, E))$ player (a) can gain in the game. Moreover, it was declared that $\mu(\Gamma(G)) \geq \zeta(G)$.

We can also describe ζ_0 recursively in the following way:

$$\zeta_0(G(V, E)) = \begin{cases} 0, & \text{if } G(V, E) \text{ is empty} \\ \infty, & \text{if } G(V, E) \text{ has an isolated vertex} \\ \max_{e \in E} \min(\zeta_0(G(V, E) - e), \zeta_0(G(V, E) - \Delta[x] - \Delta[y]) + 1), & \text{otherwise.} \end{cases}$$

Aharoni and colleagues [1] proved the following assumption:

Theorem 3.

$$\mu(\Gamma(G(V, E))) = \zeta_0(G(V, E))$$

Kawamura [2] showed the theorem for chordal graphs. Now, we describe a chordal graph that is defined as follows: A graph is chordal if each of its cycles of four or more nodes has a chord, which is an edge linking two nodes that are not together in the cycle.

The chordal graphs are the joint graphs of subtrees of a tree. A typical theorem of Dirac [3] declares that for each chordal graph G , there is a vertex, named a simplicial vertex, such that $\Delta(v)$ is a complete graph.

In a previous study [4], Abd Algani showed the theorem for circular-arc graphs, which is known as the joining graphs of a set of arcs on a circle. This graph comprises one vertex for each arc in the set and an edge connecting every two vertices whose arcs intersect.

Officially, let

$$\Gamma_1, \Gamma_2, \dots, \Gamma_n \subseteq S_1$$

be a set of arcs. Then, the corresponding circular-arc graph is G , such that

$$V = \{\Gamma_1, \Gamma_2, \dots, \Gamma_n\}$$

and

$$\{\Gamma_\alpha, \Gamma_\beta\} \in E \Leftrightarrow \Gamma_\alpha \cap \Gamma_\beta \neq \emptyset.$$

A group of edges that parallels to $G(V, E)$ is termed an arc model of $G(V, E)$.

In [5], the following theorem was demonstrated.

Theorem 4. Let G be a *min* counterexample for Theorem 3. Then, G has no vertex of degree 1.

Moreover, Theorems 1 and 2 grasp for each tree.

In addition, Berger [5] proved the following:

Theorem 5. Let $G(V, E)$ be a chordal graph; then, $\zeta_0(\overline{G(V, E)}) = \infty$.

Theorems 2, 3, and 5 were shown in [5] for complements of chordal graphs and for trees. Kawamura [2] verified the theorem for chordal graphs.

Theorem 6. Let $G(V, E)$ be a chordal graph. Then, $\Gamma(G(V, E))$ is either contractible¹ or is homotopy² equal to the section of finitely several spheres $\bigvee S^{k_t}$ where $k_t \geq \delta(G(V, E)) - 1$ for each k_t .

On the contrary, all limited wedges appear as homotopy types of independence complex of chordal graphs.

Kawamura [2] proves the following theorem:

Theorem 7. Let L be a simplicial complex and let k_1, \dots, k_r be subcomplexes of L (repetitions allowable). Consider individual points u_1, \dots, u_r and v with $\{u_1, \dots, u_r, v\} \cap L$; then, the union

$$\Lambda = \mathbb{C}_v L \cup \bigcup_{i=1}^r (C_{u_i} \cap K_i)$$

subject to the conditions

$$C_{u_i} K_i \cap C_v L = K_i \text{ and}$$

$$\mathbb{C}_{u_i} K_i \cap \mathbb{C}_{u_j} K_j = K_i \cap K_j (i, j = 1, \dots, r, i \neq j).$$

a. Is homotopy corresponding to $\bigvee_{i=1}^r \text{susp} K_i$

b. The suspension $\text{susp}(\Lambda)$ is contractible for each contractible complex Λ .

c. $\text{susp}(\bigvee S^{k_r}) \approx \bigvee S^{k_r+1}$.

Theorems 7–10 [1, 3] offer us a strong understanding of the construction of chordal graphs' independence complexes.

For a vertex v of a graph $G(V, E)$, let $\Gamma_v(G(V, E))$ be the subcomplex generated by the independent sets containing v :

$$\Gamma_v(G(V, E)) = \{U \mid \text{there exists a simplex } B \in \Gamma(G(V, E)) \text{ such that}$$

$$\{v \in B \text{ and } U \subset B\}.$$

Theorem 8. Let $G(V, E)$ be a graph and let v be a simplicial vertex of $G(V, E)$. Enumerate $\Delta(v)$ as $\Delta(v) = \{w_1, \dots, w_r\}$. Then, we have the following:

¹ Let X be a topological space; we said that X is contractible if the identity map on the topological space X is null-homotopic.

² Two continuous functions from one topological space to another are said to be homotopic if one can be “continuously deformed” into the other; this distortion is referred to as the homotopy between the two functions.

$$\Gamma(G(V, E)) \approx \bigvee \text{susp} \Gamma(G - \Delta[w_i]).$$

For every chordal graph $G(V, E)$, we term $\mu(G(V, E)) \in \mathbb{Z}^+$ as follows: if $\Gamma(G(V, E))$ is contractible, then let $\mu(G(V, E)) = \infty$, and if $\Gamma(G(V, E))$ is homotopy equal to the wedge $\bigvee S^i$, then let $\mu(G(V, E)) = \min(i_r)$, the minimum dimension of the related spheres. Theorems 6 and 7 offer the following theorem:

Theorem 9. Let $G(V, E)$ be a chordal graph; then, $\mu(G(V, E)) \geq \delta(G(V, E)) - 1$.

Theorem 10. Let v be a simplicial vertex of the graph $G(V, E)$ and enumerate $\Delta(v)$ as $\Delta(v) = \{w_1, \dots, w_r\}$. Then, we have the equation $\mu(G(V, E)) = \min\{\mu(G(V, E) - \Delta[w_k]) | k = 1, \dots, r\} + 1$.

And we can agree that

$$\min\{\infty, \dots, \infty\} + 1 = \infty.$$

For the path graph P_n with n edges, $\Gamma(P_n)$ is contractible if $n \equiv 0 \pmod{3}$ and is homotopy equal to the sphere of $\lfloor \frac{n}{3} \rfloor$ dimension. Then, we have the following equation:

$$\mu(P_n) = \begin{cases} \infty, & \text{if } n \equiv 0 \pmod{3} \\ \frac{n}{3}, & \text{otherwise.} \end{cases}$$

It should be noted that Kawamura [2] proved Theorems 6–10 for chordal graphs. In previous research [4], Abd Algani proved Theorem 2 about circular-arc graphs. In addition, in the same paper, he found that for path graph (P_n) :

$$\mu(P_n) = \zeta_0(P_n) = \begin{cases} \infty, & \text{if } n \equiv 0 \pmod{3} \\ \frac{n}{3}, & \text{otherwise.} \end{cases}$$

Moreover, for the general case of circle graph (C_n) , it was determined that

$$\mu(C_n) = \min(\mu(P_{n-5}) + 1, \mu(P_{n-6}) + 2) = \frac{n}{3}.$$

2. New result

In a previous paper, Abd Algani [4] demonstrated the following theorem:

Let $G(V, E)$ be a circular arc graph; if $\zeta_0(G(V, E)) \leq 2$, then $\mu(\Gamma(G(V, E))) \leq 2$.

In this chapter, the theorem will be generalized for a bipartite graph.

Main theorem: Let G be a bipartite graph; if $\zeta_0(G(V, E)) \leq 2$, then $\mu(\Gamma(G(V, E))) \leq 2$.

Proof.

Let $G(V, E)$ be a bipartite graph, that is, a graph with the property: if we remove each edge in the graph with all its neighbors, the graph that we will obtain is a complete bipartite graph.³

We need to prove that if $\zeta_0(G) \leq 2$, then $\mu(\Gamma(G)) \leq 2$.

³ A complete bipartite graph is a bipartite graph that contains all possible edges.

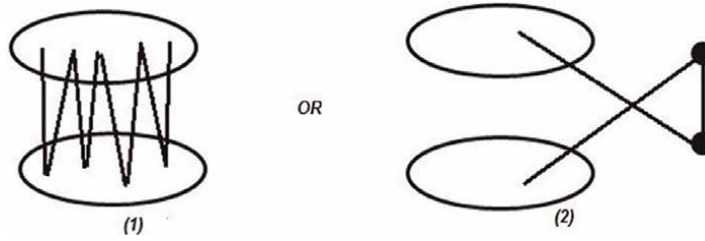


Figure 1.
The graph looks like (1) or (2) when we remove all possible edges.

First, we remove all the possible edges. In this situation, the graph will look like **Figure 1**.

The graph of \mathbb{C}_6 is a bipartite graph, and each set is an independent set (see **Figure 2**). Now, color the sets by two colors, for example, red and blue, as follows: Mark the graph vertices with colors such that no two vertices sharing the same edge have the same color (see **Figure 2**).

Each set has a vertex, where the set front does not have a neighbor, so we have built the largest set of vertices (see **Figure 2**).

The result leads to a construction, as clarified in **Figure 3**.

The result of performing a retraction⁴ is a cube. The cube has two foreign edges and two various vertices (see **Figure 4**). If the path P_4 exist in the bipartite graph and the path P_5 not exist, then:

If we insert a vertex into the path P_4 , as clarified in **Figure 5**: If we remove their edge 1 (see **Figure 5**), then we have two isolated vertices.

If we have path P_4 and path P_5 and do not have \mathbb{C}_6 (see **Figure 6**). In **Figure 6**, if we remove edge $\{3, 4\}$ with all of its neighbors, the remaining graph is a complete bipartite graph; then, vertex (1) and vertex (6) remain and are connected by an edge, and we obtain graph \mathbb{C}_6 .

If we cannot remove the edges from the graph, then $\mu = 2$.

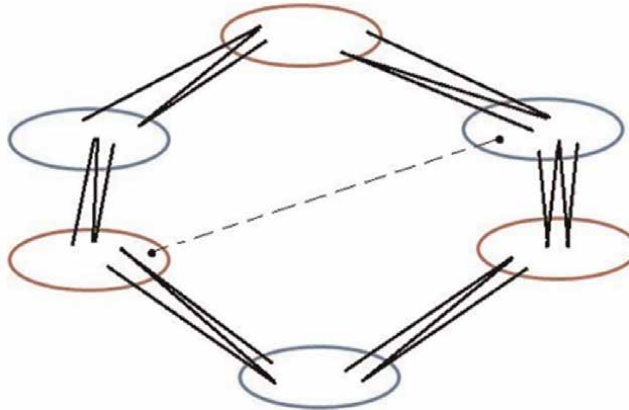


Figure 2.
The inflated graph of \mathbb{C}_6 .

⁴ Definition: Let X be a topological space, let $A \subseteq X$, and let $i : A \hookrightarrow X$ be the inclusion map. Continuous map $f : X \rightarrow A$ is called a retraction iff $f|_A = id_A$.

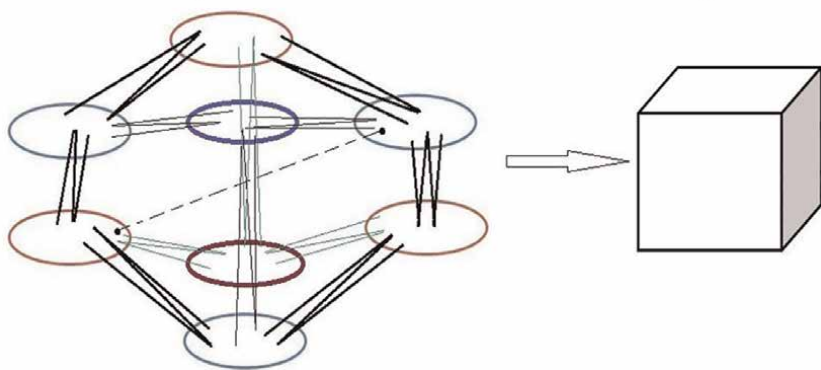


Figure 3.
The result of performing a retraction is a cube.

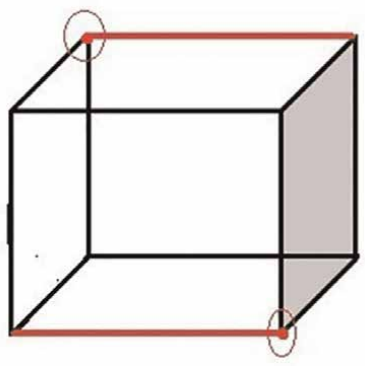


Figure 4.
The cube has two foreign edges and two various vertices.

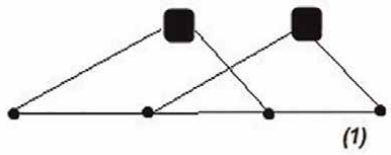


Figure 5.
The form of the neighbors of P_4 .

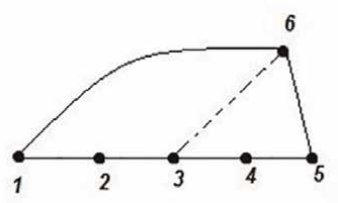


Figure 6.
Without C_6 .

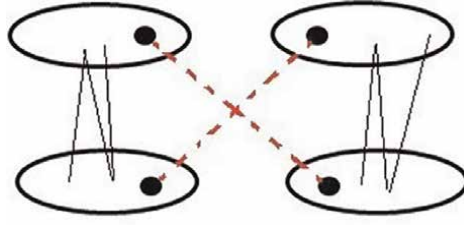


Figure 7.

If there are no red edges, then we perform a retraction, with the result being an induced matching.

Now, we prove that $\mu = 2$ before we remove any edges. We have three cases.

Case 1: The graph resembles **Figure 7**; the red edges may not exist. If there are no edges (red edges in **Figure 7**), then we perform a retraction, with the result being an induced matching.

Case 2: With the existence of the red edges, the graph will resemble **Figure 8**: If we remove the “Blue” edge (between two sets) with all of its neighbors, then the vertex V_1 will be isolated.

We can remove the “Green” edges in **Figure 8**, and we can remove the (main) “Black” edges in the same figure, then we have two connected components.

Note that the situation, as described in **Figure 9**, cannot exist, because there is a vertex in set 1, which is not connected to a vertex in set 3 by an edge, and therefore, the complementary complex is a chordal graph; thus, $(\mu = \zeta_0 = \infty)$.

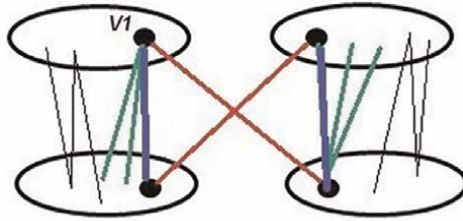


Figure 8.

Existence of red edges.

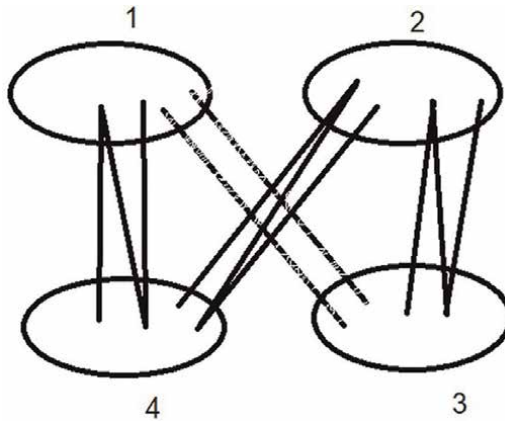


Figure 9.

This situation cannot exist.

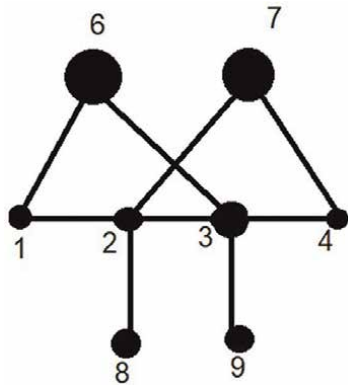


Figure 10.
 If we remove edge $\{3, 4\}$, then vertex V_1 is isolated.

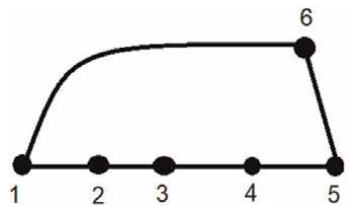


Figure 11.
 We cannot remove edge $\{1, 2\}$.

If we have a bipartite graph, and we know that the complementary independent complex of a bipartite graph is a chordal graph, then $\mu = \zeta_0 = \infty$.

In the case of P_4 (see **Figure 10**), if we remove edge $\{3, 4\}$, then the vertex V_1 is isolated.

We cannot remove edge $\{2, 3\}$ because the vertices v_2 and v_3 are two dominating vertices.

Case 3: After we remove all possible edges, we obtain P_5 (but do not have C_5); we cannot remove edge $\{1, 2\}$ (see **Figure 11**) because v_6 is a neighbor of v_5 . This situation is C_6 .

Thus, we have found the independence complex of the topological connectivity of the bipartite graph:

if

$$\zeta_0(G) \leq 2$$

then

$$\mu(c_i(G)) \leq 2.$$

3. Conclusion

In this chapter, we found that the topological connectivity of the bipartite graph's independence complex where the theorem that if $\zeta_0(G(V, E)) \leq 2$, then $\mu(\Gamma(G(V, E))) \leq 2$ is proven for a bipartite graph $G(V, E)$ by examining all possible cases. For future


research, we recommend trying to prove that $\mu(\Gamma(G(V, E))) \geq \zeta_0(G(V, E))$ for bipartite chordal graphs (the topological connectivity of independence complex of bipartite chordal graph). It is worth noting that the topological connectivity of independence complex has been proven previously for trees and complements of chordal graphs by Berger [5], for chordal graphs by Kawamura [2], and for circular-arc graph by Abd Algani [4].

Author details

Yousef Methkal Abd Algani* and Amal Sharif-Rasslan
The Arab Academic College for Education in Israel, Haifa, Israel

*Address all correspondence to: yosefabdalgani@arabcol.ac.il

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Aharoni R, Berger E, Ziv R. Independent systems of representatives in weighted graphs. *Combinatorica*. 2007;27:253-267. DOI: doi.org/10.1007/s00493-007-2086-y
- [2] Kawamura K. Independence complex of chordal graphs. *Elsevier, Discrete Mathematics*. 2010;310:2204-2211. DOI: doi.org/10.1016/j.disc.2010.04.021
- [3] Dirac GA. On rigid circuit graphs. *Abhandlungen aus dem Mathematischen Seminar der Universität Hamburg*. 1961; 25:71-76. DOI: doi.org/10.1007/BF02992776
- [4] Abd AY. The topological connectivity of the Independence complex of circular-arc graphs. *Universal Journal of Mathematics and Applications*. 2019; 2(4):159-169. DOI: [DOI.org/10.32323/ujma.556457](https://doi.org/10.32323/ujma.556457)
- [5] Berger E. Topological Methods in Matching Theory [thesis]. Faculty Of Princeton University in Candidacy; 2004

Section 4

Applications of Topology
in Physics

Anomalies and Their Occurrence in the Study of Topology in Yang-Mills Gauge Theories

Paul Bracken

Abstract

When an exact symmetry of the classical action is not preserved as a symmetry under quantization, such as path integral quantization, a gauge theory such as Yang-Mills is said to be anomalous. It is the intention to introduce the subject of anomalies and study briefly some of the topological implications that are associated with the presence of non-abelian gauge anomalies in quantum field theories.

Keywords: anomaly, topology, gauge, group, manifold, connection, bundle, fiber, path integral

1. Introduction

The objective is to study the topological and cohomological properties of abelian and non-abelian chiral anomalies, which arise in the context of Yang-Mills gauge theories [1–3]. The occurrence of anomalies can be traced back to the nontrivial topology of the configuration or Yang-Mills orbit space. The abelian chiral anomaly in $d = 2p$ dimensions and the non-abelian gauge anomalies when $d = 2$ can be calculated. These results may be interpreted in terms of suitable index theorems on spaces of dimension $d = 2p$ and $d + 2$, respectively. There are consistency conditions for the anomalies and the Schwinger terms can be interpreted in terms of cohomology such that the cocycles take values in the space of functionals of the gauge fields. The cohomological descent procedure which starts from the Chern character forms provides a method for obtaining nontrivial candidates for both the non-abelian anomalies and the Schwinger terms.

The requirement that a field theory be invariant under rigid transformations $\{g_i\}$ of a gauge group G as well as be invariant under local $\{g_i = g_i(x)\}$ transformations is called the gauge invariance principle. This symmetry principle is dynamical because besides including rigid symmetry, it also yields information about the way the interaction with the gauge fields has to be expressed. This gauge invariance is obtained by introducing gauge or Yang-Mills potentials $A_\mu^j(x)$, such that j is a group index. These transform under the adjoint representation of a compact Lie group G and $x = (x^\mu)$ is the spacetime manifold coordinate. The coordinate derivative ∂_μ is gauged by the covariant derivative D_μ . The Yang-Mills potentials are subject to gauge

transformations $A \rightarrow A^g$, so independent degrees of freedom have to be identified to avoid overcounting. These symmetries express redundancy in the description of the Yang-Mills fields. Let \mathcal{A} be the infinite-dimensional space of all Yang-Mills potentials A . The action of the gauge group of \mathcal{A} determines orbits, which contain the Yang-Mills field connected to it by a gauge transformation. These extra degrees of freedom can be eliminated by fixing a gauge. This means finding a solution $A_\mu^{g_0}$ to a gauge fixing constraint $f(A^g) = 0$. The $A_\mu^{g_0}$ of a given orbit is the potential satisfying the same condition as f . In non-abelian theories the solution to this equation is not unique. For example, the Coulomb gauge condition does not determine A uniquely, since $\nabla A' = \nabla A$ allows these potentials to be related, that is, when $\Delta \tau = 0$, $A' = A + \nabla \tau$. Thus, there is no unique intersection between the surface $f(A_\mu^g) = 0$ and the orbits A_μ^g . The phenomenon known as the Gribov ambiguity implies no unique intersection between the surface $f(A^g) = 0$ and the orbits [4]. However, as the intersections of an orbit with the surface $f(A^g) = 0$ are separated by finite gauge transformations, this ambiguity does not disturb the perturbative representation of the theory, which is based on an expansion of a classical configuration. The geometrical nature behind the ambiguity is that of a nontrivial principle bundle P , the bundle of Yang-Mills potentials for which there is no global section. The structure group of this bundle is the group of gauge transformations. The fibers are the orbits of the potentials and the base is determined by independent gauge degrees of freedom. Selecting a gauge would correspond to choosing a vector potential in each orbit in a continuous way. No global gauge fixing is possible on compactified spacetimes. In trying to extend the local chart to the whole manifold, the effect is that, beyond a certain distance, the gauge condition does not fix the gauge uniquely. There are better ways of thinking about a gauge transformation. For example, it is possible to look at a gauge transformation as the change produced in A_μ by a vertical bundle automorphism f of P , where f is trivial on the base space $f_b = 1_M$. If $f : P \rightarrow P$ is an automorphism $f^*(\omega)$ is the gauge transformed connection. So, when $A = \sigma^*(\omega)$ on U , then the new A' is given by $\sigma^*(f^*\omega) = (f \circ \sigma)^*(\omega) = A'$, where in analogy we denote $f \circ g$ by σ' and both are equivalent.

2. Geometry of gauge theories

Let us provide more information about gauge theories that are required along the way. The group of gauge transformations on spacetime M is the group $\mathcal{C} = \text{Aut}_V(P)$ of vertical automorphisms of $P(G, M)$, or $\text{Aut}_V P = \{f \in \text{Diff}_M(P) | f_b = I_M\}$, so f is trivial on the base. A Yang-Mills potential comes from a connection ω on a principle bundle $P(G, M)$ over a spacetime M such that gauge transformation $g(x)$ relates the local representatives $\sigma^*(\omega) = A$ and $A' = (f \circ \sigma)^*(\omega)$ determined by the pullback of ω by two local sections σ and σ' over $U \subset M$. If $\sigma_\alpha^*(\omega) = A_\alpha$ and $\sigma_\alpha'(x) = f(\sigma_\alpha(x)) = \sigma_\alpha(x)\varphi(\sigma_\alpha(x)) = \sigma_\alpha\varphi_\alpha(x)$ the gauge transformation is

$$A'_\alpha(x) = \text{Ad}(\varphi_\alpha^{-1}(x))A_\alpha(x) + \varphi_\alpha^{-1}(x)d\varphi_\alpha(x). \quad (1)$$

If two local sections are given $\sigma_\beta(x) = \sigma_\alpha(x)g_{\alpha\beta}(x)$, (then)

$$\sigma'_\beta(x) = f(\sigma_\beta(x)) = \sigma_\beta(x)\varphi_\beta(x) = f(\sigma_\alpha(x)g_{\alpha\beta}(x)) = \sigma_\alpha(x)\varphi_\alpha(x)g_{\alpha\beta}(x). \quad (2)$$

Consequently, we express $\varphi_\beta = g_{\alpha\beta}^{-1}\varphi_\alpha(x)g_{\alpha\beta}(x)$ for $x \in U_\alpha \cap U_\beta$. The elements of the gauge group \mathcal{C} of gauge transformations are sections in $\Gamma(\tilde{\text{Ad}}(P), M)$. It can be shown that \mathcal{C} is a Baker-Campbell-Hausdorff group. The Lie algebra is given by sections of the associated bundle of Lie algebras $\text{Ad}(P) = P \times_{\text{Ad } G} \mathcal{C}$. Thus, a gauge transformation can be viewed as a section of $\tilde{\text{Ad}}(P)$, and an infinitesimal gauge transformation is a section of $\text{Ad}(P)$ [5–7].

Gauge-unrelated Yang-Mills potentials are projected onto different points of the quotient \mathcal{A}/\mathcal{C} , the orbit space of the Yang-Mills connections on $P(M, G)$. In general, the action of \mathcal{C} on \mathcal{A} is not free but with certain technical restrictions, it is possible to obtain a free action. This may be done, for example, by restricting \mathcal{C} to the group of automorphisms, preserving a point $p \in P$ or to the group of based gauge transformations, that is, leaving infinity or a point of $M^m = S^m$ fixed. It is not hard to see the bundle of connections is not trivial in general. If it were, it would be permissible to write $\mathcal{A} = \mathcal{C} \times \mathcal{A}/\mathcal{C}$. The functional space of connections \mathcal{A} is an affine space, and hence, it is contractible and it has no topological invariants. However, $\mathcal{A} = \mathcal{C} \times \mathcal{A}/\mathcal{C}$ implies that $0 = \pi_j(\mathcal{A}) = \pi_j(\mathcal{C}) + \pi_j(\mathcal{A}/\mathcal{C})$ this cannot be fulfilled in general since \mathcal{C} possesses non-zero homotopy groups. The bundle may be nontrivial and no continuous gauge fixing is possible. Since \mathcal{A} is topologically trivial, the topology of \mathcal{A}/\mathcal{C} comes entirely from \mathcal{C} .

Consider briefly the topology of the orbit space \mathcal{A}/\mathcal{C} , the space of measurable, physical fields, that is, configuration space of the Yang-Mills theory. This depends on the topology of the manifold M . The usual asymptotic conditions allow us to replace the m -dimensional space M by its conformal compactification S^m , which is useful for topological considerations. To obtain a finite Yang-Mills action, assume $A(x)$ become pure gauge at infinity $A(x) \rightarrow g^{-1}(x)dg(x)$, where $g : S^{m-1} \rightarrow G$ and S^{m-1} is the sphere at infinity. These mappings fall into homotopy classes $\pi_{m-1}(G)$, which classify the bundle $P_k(G, S^m)$ over S^m . If m is even, the class k is obtained by computing the Chern class $c_{(m/2)}$ over S^m . This k -dependence is reflected in other constructions. Thus, \mathcal{A} splits into spaces \mathcal{A}_k of connections on the bundle $P_k(G, S^m)$ with gauge group \mathcal{C}_k . For a given k , $\mathcal{A}_k(\mathcal{C}_k, \mathcal{A}_k/\mathcal{C}_k)$ is a principle bundle with structure group \mathcal{C}_k . As pointed out by Atiyah and Jones [8], the homotopy type of $\mathcal{A}_k/\mathcal{C}_k$ does not make reference to k , so $\pi_i(\mathcal{A}_k/\mathcal{C}_k) = \pi_i(\mathcal{A}/\mathcal{C})$ so we may set $k = 0$. As far as homotopy properties are concerned, the trivial bundle may be taken as $P_0(G, S^m)$ for which $P_0 = G \times S^m$, so that $\mathcal{C}_0 = G(S^m)$ and k may be ignored. From the homotopy sequence induced by $\mathcal{A}(\mathcal{C}, \mathcal{A}/\mathcal{C})$, we find that

$$0 = \pi_n(\mathcal{A}) = \pi_n(\mathcal{A}/\mathcal{C}) \rightarrow \pi_{n-1}(\mathcal{C}) \rightarrow \pi_{n-1}(\mathcal{A}) = 0, \quad (3)$$

from which we infer,

$$\pi_j(\mathcal{A}/\mathcal{C}) = \pi_j(\mathcal{C}). \quad (4)$$

The non-triviality of the orbit space is tied to the non-triviality of the homotopy groups of G . For an abelian gauge theory in four dimensions $\pi_j(\mathcal{A}/\mathcal{C}) = \pi_{j+1}(U(1)) = 0$. This means $U(1)$ -gauge bundle of quantum electrodynamics is trivial and so no Gribov ambiguity in quantum electrodynamics. The non-abelian anomaly of chiral gauge theories in four dimensions may be exposed topologically by looking for non-contractible two-spheres in orbit spaces for which $\pi_2(\mathcal{A}/\mathcal{C}) = \pi_5(G) = \mathbb{Z}$. In general, in even dimension, a sufficient condition for the existence of a perturbative non-abelian anomaly is that $\pi_2(\mathcal{A}/\mathcal{C}) = \pi_1(\mathcal{C}) = \pi_{2p+1}(G) = \mathbb{Z}$.

3. Quantization of gauge theories and the appearance of anomalies

In general, a quantum theory is called anomalous if an exact symmetry of the classical action is not preserved as a symmetry as a result of path integral quantization of the theory. It is then said there is an obstruction to the lifting of the classical symmetry in the quantum theory [9–10].

An example of this is the axial vector Noether current in a theory with massless fermions associated with the rigid chiral symmetry of the classical action that is not conserved in the quantum case. The anomaly contributes to physical processes such as the decay of the π^0 boson to two photons. When anomalies effect the gauge symmetries of the theory, it becomes inconsistent. To have unitarity and renormalizability in $d = 4$, gauge invariance turns out to be a crucial property of Yang-Mills theories [11].

A brief discussion of the $U(1)$ anomaly, or chiral anomaly, will be presented in an arbitrary d -dimensional space. Some of the explicit calculations have been suppressed, so we can concentrate on studying topology. The fact d is even is due to the fact that chirality can only be defined in even dimensions. One way to proceed is to use the mathematical approach of Fujikawa [12], which leads ultimately to a field-theoretic illustration of the index theorem. In classical theories, all that matters is the Lagrangian. In the quantum case, it is the nonlocal functional $W[A]$ that is the relevant quantity

$$W[A] = \int \mathcal{D}\psi \mathcal{D}\bar{\psi} e^{iS[A, \psi]} = e^{iZ[A]}. \quad (5)$$

In (5) $\mathcal{D}\psi \mathcal{D}\bar{\psi}$ is the fermionic functional integration measure and the action S is given as

$$S[A, \psi] = \int dx \bar{\psi} i \not{D} \psi, \quad D_\mu = \partial_\mu + A_\mu, \quad A_\mu = A_\mu^i T_i, \quad (6)$$

with $\not{D} = \gamma^\mu D_\mu$ and $\psi, \bar{\psi}$ are independent Grassmann variables. The integration measure in d dimensions is written simply as $d\mathbf{x}$. The T_i are the generators of the Lie algebra in matrix form, act on the group representation index in ψ . The Minkowski metric on spacetime is $\eta^{\mu\nu} = (+, -, \dots, -)$ and the gamma matrices satisfy $\{\gamma^\mu, \gamma^\nu\} = 2\eta^{\mu\nu}$, $\gamma^{0\dagger} = \gamma^0$, $\gamma^{i\dagger} = -\gamma^i$, for $i = 1, 2, \dots, (d-1)$ and $\gamma_{d+1} = (i)^{d/2-1} \gamma^0 \dots \gamma^{d-1}$. The exponential in (5) has an oscillatory nature, so the path integral is not well defined because of this. This problem can be overcome by moving it to Euclidean space

$$W_E[A] = \int \mathcal{D}\psi \mathcal{D}\bar{\psi} e^{-S_E(A, \psi)} = e^{-Z_E[A]}. \quad (7)$$

Here, $S_E(A, \psi)$ is the Euclidean action and the Dirac matrices entering \not{D} are all hermitian now: $\{\gamma^\mu, \gamma^\nu\} = 2\delta^{\mu\nu}$, $\gamma^{\mu\dagger} = \gamma^\mu$, $\gamma_{d+1} = i^{d/2} \gamma^0 \dots \gamma^{d-1}$, $\gamma_{d+1}^2 = 1$, $\gamma_{d+1}^\dagger = \gamma_{d+1}$, $\gamma_{d+1} \gamma^\mu = -\gamma^\mu \gamma_{d+1}$. The rotation that changes $W[A]$ into $W_E[A]$ is given by the substitutions $x_\mu^0 = -ix_E^0$, $x_M^i \rightarrow x_E^i = x_{Ei}$, $A_0^M \rightarrow iA_0^E$, $A_i^M \rightarrow A_i^E = A^{Ei}$, $i = 1, \dots, (d-1)$ and $\psi_M \rightarrow \psi_E$, $\bar{\psi} \rightarrow i\bar{\psi}_E$, and $\gamma_M^i = i\gamma_E^i$, $\gamma_M^0 = \gamma_E^0$ are identified.

The system is invariant under rigid chiral transformations of the Dirac fields $\psi' = e^{i\alpha \gamma_{d+1}} \psi$, $\delta\psi = i\alpha \gamma_{d+1} \psi$. Classically, this means there is a conserved Noether chiral current or axial current j_{d+1}^μ , which can be deduced by writing the variation of $S[A, \psi]$ under a nonconstant parameter transformation $\alpha = \alpha(x)$ as

$$\delta S = \int d\mathbf{x} \partial_\mu \alpha j_{d+1}^\mu, \quad j_{d+1}^\mu = \bar{\psi} \gamma^\mu \gamma_{d+1} \psi. \quad (8)$$

Quantum mechanically, the axial vector current is no longer conserved. There is an anomalous divergence

$$\partial_\mu J_{d+1}^\mu = \frac{1}{W[A]} \int \mathcal{D}\psi \mathcal{D}\bar{\psi} \partial_\mu j_{d+1}^\mu e^{iS[A,\psi]}. \quad (9)$$

The quantity $\partial_\mu J_{d+1}^\mu$ is called the abelian or axial anomaly. Abelian refers to the fact that symmetry is an axial $U(1)$ transformation. It has been pointed out that the nonconservation of axial current j_{d+1}^μ may be connected to the transformation properties of the functional fermionic measure in the quantum case.

Suppose we try to derive current conservation by using the rigid chiral transformations of the Dirac fields above with $\alpha = \alpha(x)$. It is required the path integral retain the same form under this change of variable. Expanding the difference between the two integrals in terms of α , one before and one after the change using (3.4)

$$0 = \int \delta(\mathcal{D}\psi \mathcal{D}\bar{\psi}) e^{-S_E[A,\psi]} + \int \mathcal{D}\psi \mathcal{D}\bar{\psi} e^{-S_E[A,\psi]} \cdot \int d\mathbf{x} (\partial_\mu \alpha) j_{d+1,E}^\mu. \quad (10)$$

Were the fermionic measure invariant, the second term should go to zero or $\partial_\mu J_{d+1,E}^\mu = 0$. However, the path integral measure is not invariant. The value of $\delta(\mathcal{D}\psi \mathcal{D}\bar{\psi})$ in (10) is minus twice the trace, in the sense of the space of functions, of $i\alpha \gamma_{d+1}$,

$$D\psi' D\bar{\psi}' = \exp(-2 \text{Tr}(i\alpha \gamma_{d+1})) D\psi D\bar{\psi}. \quad (11)$$

What appears to be the inverse of the usual Jacobian occurs because Grassmann odd variables are involved. If Euclidean space is compactified to the sphere S^d , both ψ and $\bar{\psi}$ can be expanded in terms of the eigenfunctions of i/D with real eigenvalues λ_n

$$\psi = \sum_n a_n \psi_n, \quad \bar{\psi} = \sum_n \bar{b}_n \bar{\psi}_n, \quad (\psi_i, \psi_j) = \int d\mathbf{x} \psi_i^\dagger \psi_j = \delta_{ij}. \quad (12)$$

The character of both ψ and $\bar{\psi}$ is carried by the coefficients a_n, \bar{b}_n . The operator i/D , which is elliptic in Euclidean space, is a Fredholm operator, which means it has a discrete spectrum. The Jacobian is $2 \int d\mathbf{x} \sum_n \alpha \psi_n^\dagger i \gamma_{d+1} \psi_n$. Therefore, (10) is

$$0 = \int D\psi D\bar{\psi} \left[-2 \int d\mathbf{x} \alpha \sum_n \psi_n^\dagger i \gamma_{d+1} \psi_n - \int d\mathbf{x} \partial_\mu j_{d+1,E}^\mu \alpha \right] e^{-S_E[A,\psi]}. \quad (13)$$

This holds for any $\alpha(x)$, so from (9), we can write

$$\partial_\mu J_{d+1,E}^\mu = \frac{1}{W_E[A]} \int \mathcal{D}\psi \mathcal{D}\bar{\psi} \partial_\mu j_{d+1,E}^\mu e^{-S_E[A,\psi]} = -2i \sum_n \psi_n^\dagger \gamma_{d+1} \psi_n. \quad (14)$$

Therefore, the anomalous divergence of the current is a consequence of the nontrivial Jacobian associated with the transformation.

Expression (14) is divergent but can be regularized using the standard heat kernel regularization procedure. Therefore, calling (14) Q_E , we have

$$Q_E = \lim_{M \rightarrow \infty} (-2i) \sum_n e^{-\lambda_n^2/M^2} \psi_n^\dagger \gamma_{d+1} \psi_n. \quad (15)$$

Suppose the integral $\int d\mathbf{x} Q_E$ is considered now. Then, the ψ_n corresponding to zero modes makes a contribution to the integral. If $i/D\psi_n = \lambda_n \psi_n$, then $i/D\gamma_{d+1}\psi_n = -\lambda_n(\gamma_{d+1}\psi_n)$. For $\lambda_n \neq 0$, ψ_n and $(\gamma_{d+1}\psi_n)$ are eigenvectors that correspond to different eigenvalues of a hermitian operator, which means they are orthogonal. Introducing $P_+ = (1 + \gamma_{d+1})/2$ and $P_- = (1 - \gamma_{d+1})/2$, $\gamma_{d+1} = P_+ - P_-$. This is just the difference between the positive and negative chirality zero modes. Hence, they are related to the index of the operator in this way

$$\int d\mathbf{x} Q_E = -2i [\dim \ker(i/D_+) - \dim \ker(i/D_-)] = -2i \operatorname{ind}(i/D_+). \quad (16)$$

The elliptic non-self adjoint $(i/D)^\dagger = i/D_-$ operator for each chirality are

$$i/D_\pm = i/D P_\pm, \quad i/D = i/D_+ + i/D_-. \quad (17)$$

Thus, the integration of the abelian anomaly gives the index for the twisted spin complex associated with the operator i/D_+ . It is possible to compute (15) directly by writing it as

$$\begin{aligned} Q_E &= -2i \lim_{M \rightarrow \infty} \lim_{x \rightarrow y} \sum_n \psi_n^\dagger(y) \gamma_{d+1} e^{-(\mathcal{D})^2(x)/M^2} \psi_n(x) \\ &= -2i \lim_{M \rightarrow \infty} \lim_{x \rightarrow y} \operatorname{Tr} \left(\gamma_{d+1} e^{-(\mathcal{D})^2(x)/M^2} \sum_n \psi_n(x) \psi_n^\dagger(y) \right) \\ &= 2i \lim_{M \rightarrow \infty} \lim_{x \rightarrow y} \operatorname{Tr} \left(\gamma_{d+1} e^{-(\mathcal{D})^2(x)/M^2} \right) \delta^d(x - y). \end{aligned} \quad (18)$$

In the last step, the completeness relation $\sum_n \psi_n(x) \psi_n^\dagger(y) = \delta^d(x - y)$ for ψ_n was used. The trace involves both the group and spinorial indices. Expressing $\delta^d(x - y)$ with respect to a plane wave basis,

$$\delta^d(x - y) = \frac{1}{(2\pi)^d} \int d\mathbf{k} e^{ik \cdot (x - y)}. \quad (19)$$

Thus, for Q_E , we obtain

$$Q_E = -2i \lim_{M \rightarrow \infty} \frac{1}{(2\pi)^d} \int d\mathbf{k} e^{-ik \cdot x} \operatorname{Tr} \left(\gamma_{d+1} e^{-(\mathcal{D})^2/M^2} \right) e^{ik \cdot x}. \quad (20)$$

It is now required to deal with the limit $M \rightarrow \infty$. It suffices to know the terms that contribute to this limit. Since

$$\begin{aligned} /D^2 &= D_\mu D_\nu \gamma^\mu \gamma^\nu = \frac{1}{2} D_\mu D_\nu \{\gamma^\mu, \gamma^\nu\} + \frac{1}{2} D_\mu D_\nu [\gamma^\mu, \gamma^\nu] = D^\mu D_\mu + \frac{1}{4} [D_\mu, D_\nu] [\gamma^\mu, \gamma^\nu] \\ &= D^\mu D_\mu + \frac{1}{4} F_{\mu\nu} [\gamma^\mu, \gamma^\nu]. \end{aligned} \quad (21)$$

Introduce a new variable s defined in terms of k as $s^\mu = k^\mu / M$ in which case (15) becomes

$$Q_E = -2i \lim_{M \rightarrow \infty} \frac{1}{(2\pi)^d} \int d\mathbf{x} M^d \text{Tr} \left(\gamma_{d+1} \exp \left(- \left(\frac{i}{M} / D - /s \right)^2 \right) \right). \quad (22)$$

The exponential term can be reexpressed using (21),

$$\exp \left(- \left(\frac{i}{M} / D - /s \right)^2 \right) = \exp \left(\frac{D^2}{M^2} + \frac{1}{4M^2} F_{\mu\nu} [\gamma^\mu, \gamma^\nu] + \frac{2i}{M} s^\mu D_\mu \right) e^{-s^2}. \quad (23)$$

It is the case that all potentially divergent terms (22) vanish since $\text{Tr} (\gamma_{d+1} \gamma^{\mu_1} \dots \gamma^{\mu_k})$ for $k < d$. The result of the limit is the contribution to the exponential (23) proportional to $1/M^d$ and at the same time has enough gamma matrices. It is one that arises entirely from the middle term in (23), $F_{\mu\nu} [\gamma^\mu, \gamma^\nu]$, so Q_E is calculated

$$\begin{aligned} Q_E &= \frac{1}{(d/2)!} \frac{-2i}{(4\pi)^d} \int d\mathbf{x} e^{-s^2} \text{Tr} (\gamma_{d+1} F_{\mu_1 \mu_2} \dots F_{\mu_{d-1} \mu_d} [\gamma^{\mu_1}, \gamma^{\mu_2}] \dots [\gamma^{\mu_{d-1}}, \gamma^{\mu_d}]) \\ &\quad - \frac{2i}{(4\pi)^d} \frac{1}{(d/2)!} (2\pi)^{d/2} \text{Tr} (F_{\mu_1 \mu_2} \dots F_{\mu_{d-1} \mu_d}) \text{Tr} (\gamma_{d+1} \gamma^{\mu_1} \dots \gamma^{\mu_d}) \\ &= -2i \left(\frac{-i}{4\pi} \right)^{d/2} \frac{1}{(d/2)!} \epsilon^{\mu_1 \dots \mu_d} \text{Tr} (F_{\mu_1 \mu_2} \dots F_{\mu_{d-1} \mu_d}). \end{aligned} \quad (24)$$

The d -dimensional integral $\int d\mathbf{s} e^{-s^2} = \pi^{d/2}$ has been substituted as well as the trace equation $\text{Tr} (\gamma_{d+1} \gamma^{\mu_1} \dots \gamma^{\mu_d}) = (-2i)^{d/2} \epsilon^{\mu_1 \dots \mu_d}$.

To obtain the anomaly in d -dimensional Minkowski spacetime, a rotation from Euclidean space back is carried out. There is a factor $(-i)$ coming from the trace part $(\partial_0^E \rightarrow -i\partial_0^M)$, $(A_0^E \rightarrow -iA_0^M)$, a factor $(-1)^{d/2}$ since $(\partial_\mu J_{d+1}^\mu)_E$ goes into $(-1)^{d/2} \partial_\mu J_{d+1}^\mu$ and a minus sign, since in Minkowski spacetime $\epsilon^{0 \dots d-1} = (-1)^{d-1}$. The anomaly in Minkowski spacetime is found to be

$$Q = \left(\frac{i}{4\pi} \right)^{d/2} \frac{2}{(d/2)!} \epsilon^{\mu_1 \dots \mu_d} \text{Tr} (F_{\mu_1 \mu_2} \dots F_{\mu_{d-1} \mu_d}). \quad (25)$$

This is what results from the Feynman triangle diagram calculation when $d = 4$ for the anomaly. The abelian anomaly is gauge invariant. The last thing to do is get the index of the operator i/D_+ . Using (15), (24), and the definition of the Chern characteristic at the end, the index is calculated as

$$\begin{aligned} \text{ind } i/D_+ &= \frac{i}{2} \int d\mathbf{x} Q_E = \left(-\frac{i}{2\pi} \right)^{d/2} \frac{2}{(d/2)!} \frac{1}{2^{d/2}} \int d\mathbf{x} \epsilon^{\mu_1 \dots \mu_d} \text{Tr} (F_{\mu_1 \mu_2} \dots F_{\mu_{d-1} \mu_d}) \\ &= \left(-\frac{i}{2\pi} \right)^{d/2} \frac{1}{(d/2)!} \frac{1}{2^{d/2}} \int_{S^d} dx^{\mu_1} \wedge \dots \wedge dx^{\mu_d} \text{Tr} (F_{\mu_1 \mu_2} \dots F_{\mu_{d-1} \mu_d}) \\ &= (-1)^{d/2} \int_{S^d} \left(\frac{i}{2\pi} \right)^{d/2} \frac{1}{(d/2)!} \text{Tr} (F^{d/2}) = (-1)^{d/2} \int_{S^d} ch_{(d/2)}(F). \end{aligned} \quad (26)$$

This shows the expectation value $(\partial_\mu J_{d+1}^\mu)_E$ is given by the index of the Dirac operator. As $M \sim D^d$, the spin connection does not enter and the index density for the Dirac operator reduces to the Chern characteristic [12–14].

4. Nontrivial topology and non-abelian case

Non-abelian gauge anomalies admit path integral representation and appear in theories with gauge fields that are coupled to chiral fermions called Weyl fermions. They come up in even dimensions like the abelian anomaly. The generating function of the theory with dynamical gauge fields is made up of a kinetic term, the functional measure $\mathcal{D}A_\mu$ and so forth but there are not important for the problem of gauge invariance of the theory to be analyzed as it is concentrated in (22), since the other contributions are gauge invariant. Since gauge anomalies still make the theory inconsistent, quantize the fermion degrees of freedom first retaining A_μ as an external field to investigate the gauge transformation behavior of the resulting theory [15].

The functional is (5), but now ψ is a Weyl spinor that has positive chirality, $\gamma_{d+1}\psi = \psi$, $P_+\psi = \psi$, so \not{D} can be replaced by \not{D}_+ now. The variation of $Z[A]$ under a gauge transformation $\zeta^i(x)$ with $\zeta^i \cdot Y$ as generator is given by

$$(\zeta^i \cdot Y) Z[A] = \int d\mathbf{x} \zeta^i(x) Y_{(i)}(x) Z[A]. \quad (27)$$

The functional derivative is given by

$$\frac{\delta Z[A]}{\delta A_\mu^i(x)} = \frac{1}{W[A]} \int \mathcal{D}\psi \mathcal{D}\bar{\psi} j_{(i)}^\mu(x) \exp(iS[A, \psi]) = J_{(i)}^\mu(x), \quad (28)$$

where $j_{(i)}^\mu = \bar{\psi} \gamma^\mu (iT_{(i)}) \psi$, $(iT_{(i)})^\dagger = (iT_{(i)})$ is the current coupled to $A_\mu^i(x)$ and $J_{(i)}^\mu$ is defined by (28). Consequently, it is found that

$$(\zeta \cdot Y) Z[A] = \int d\mathbf{x} \zeta^i(x) D_\mu J_{(i)}^\mu(x) = \int d\mathbf{x} \zeta^i(x) \mathcal{V}^i[A](x). \quad (29)$$

If current $J_{(i)}^\mu$ is not covariantly conserved, a breakdown of a conservation law. The functional will not be gauge invariant and there appears an anomaly. Thus, $D_\mu j_{(i)}^\mu = 0$ does not continue to be maintained in the quantum case.

As with the abelian case, the non-invariance of $Z[A]$ may be tied to the fact that, although the exponential in (5) is gauge invariant, given by the classical action, the fermionic functional measure $\mathcal{D}\psi \mathcal{D}\bar{\psi}$ is not. This implies a Jacobian determinant appears that formally is,

$$\frac{\mathcal{D}(g^{-1}\psi) \mathcal{D}(\bar{\psi}g)}{\mathcal{D}\psi \mathcal{D}\bar{\psi}} = e^{i\alpha_1[A, g]}. \quad (30)$$

If $W[A^g] = e^{\alpha_1[A, g]} W[A]$, $W[A] = e^{iZ[A]}$, it follows that $Z[A^g] = Z[A] + \alpha_1[A, g]$. Then, $Z[A^g] - Z[A] = (\zeta \cdot Y) Z[A] = \int d\mathbf{x} \zeta^i \mathcal{V}_{(i)}[A](x)$. This means α_1 is given by

$$\alpha_1[A, g] = \int d\mathbf{x} (\zeta^i \mathcal{V}_{(i)}[A](x)) \quad (31)$$

The euclidean generating functional requires the calculation of the determinant of the matrix $i\mathcal{D}$ for its evaluation. The problem is $i\mathcal{D}$ alters the chirality here, so there is no well-defined eigenvalue problem for chiral fermion fields, hence no basis of eigenfunctions. The equation $\mathcal{D}\psi = \psi$ does not make sense for a Weyl spinor. One way to overcome this is to consider complex Dirac fermions instead of chiral Weyl fermions and replace $i\mathcal{D}$ with $i\mathcal{D}^*$ where

$$\mathcal{D}^* = \mathcal{D}P_+ + \partial\mathcal{P}_- = \partial' + \mathcal{A}_{P_+}. \quad (32)$$

The new generating functional becomes $\int \mathcal{D}\psi \mathcal{D}\bar{\psi} \exp\left(-\int d\mathbf{x} \bar{\psi} i\mathcal{D}^* \psi\right) = \det(i\mathcal{D}^*)$. With such an operator, the negative chirality fermions do not couple to A . They just give a proportionality factor in (7). Thus, the chiral gauge theory defined by \mathcal{D}^* is the same as the one based on \mathcal{D} up to this factor. This operator has a well-defined eigenvalue problem associated with it, but its eigenvalues are not real since $i\mathcal{D}$ is not a self-adjoint operator. The new action is

$$S_E^*(A, \psi) = \int d\mathbf{x} \bar{\psi} i\mathcal{D}^* \psi. \quad (33)$$

is invariant under gauge transformations $\delta\zeta = -\zeta P_+ \psi$, $\delta_\zeta \bar{\psi} = \bar{\psi} P_- \zeta$, $\delta_\zeta A_\mu = \partial_\mu \zeta + [A_\mu, \zeta]$ with $\zeta(x) = \zeta^i(x) T_i$. These are ordinary gauge transformations for the positive chirality components. Under these transformations, however, the measure is not invariant and an anomaly results. If ψ' and A' are the transformed fields

$$\begin{aligned} W_E^*[A'] &= \int \mathcal{D}\psi' \mathcal{D}\bar{\psi}' e^{-I_E^*(A', \psi')} = \int \mathcal{D}\psi' \mathcal{D}\bar{\psi}' e^{-I_E^*(A', \psi')} \\ &= \int \mathcal{D}\psi' \mathcal{D}\bar{\psi}' e^{-I_E^*(A, \psi)} = \int \mathcal{D}\psi \mathcal{D}\bar{\psi} J e^{-I_E^*(A, \psi)}. \end{aligned} \quad (34)$$

Integration variables have simply been relabelled in the second and invariance of the action used in the third equality and J is the Jacobian determinant. Since the operator $i\mathcal{D}^*$ is elliptic, it has a discrete spectrum on a compact manifold. As $i\mathcal{D}^*$ is not self-adjoint, it has eigenvalues that are complex. This means that the left and right eigenfunctions have to be introduced

$$i\mathcal{D}^* \phi_n = \lambda_n \phi_n, \quad \chi_n^\dagger i\mathcal{D}^* = \lambda_n \chi_n^\dagger, \quad \int d\mathbf{x} \chi_m^\dagger \phi_m = \delta_{mm}. \quad (35)$$

These can be used to form expansions for ψ and $\bar{\psi}$ as before $\psi = \sum_n a_n \phi_n$, $\bar{\psi} = \sum_n \bar{b}_n \chi_n^\dagger$. It is necessary to regularize the Jacobian, and so as usual, a cutoff M is introduced

$$J = \exp\left(\lim_{M \rightarrow \infty} \int d\mathbf{x} \sum_n (\chi_n^\dagger(x) \gamma_{d+1} \zeta \phi_n(x)) e^{-\lambda_n^2/M^2}\right). \quad (36)$$

To evaluate (36), it is only necessary to retain the term independent of the cutoff M . Omitting considerable calculation, the result is

$$J = \exp \left\{ \lim_{M \rightarrow \infty} \int d\mathbf{x} \lim_{x \rightarrow y} \text{Tr} \left[\gamma_{d+1} \zeta(x) e^{-(\mathcal{D})^2(x)/M^2} \delta(x-y) \right] \right\}. \quad (37)$$

and Tr indicates a trace on the spacial as well as on the gauge group indices and $\sum_n \phi_n(x) \chi_n^\dagger(y) = \delta^d(x-y)$. It can also be shown that the expression for the anomaly when $d = 4$ is

$$-\frac{i}{24\pi^2} \int \text{Tr} \left[\zeta \left(d \left(A dA + \frac{1}{2} A^2 \right) \right) \right]. \quad (38)$$

This implies that

$$(D_\mu J^\mu)_{(j)} = -\frac{i}{24\pi^2} \epsilon^{\mu\nu\rho\sigma} \text{Tr} \left[T_{(j)} \partial_\mu \left(A_\nu \partial_\mu A_\sigma + \frac{1}{2} A_\nu A_\rho A_\sigma \right) \right]. \quad (39)$$

We can now investigate the non-abelian anomaly as a probe for nontrivial topology. There is a similarity between the expression for the $U(1)$ anomaly when $d = 4$ and the non-abelian gauge anomaly. They are given by the divergence and covariant divergence of the currents J_{d+1}^μ and $J_{(i)}^\mu$. Apart from the numerical factor of the A^3 term, they are similar.

5. The anomaly as a way to the topology of gauge theories

It has been shown by Atiyah and Singer that the expression for the non-abelian anomaly in $d = 2p$ dimensions may be obtained from that of the abelian anomaly in a $(d+2)$ -dimensional space by using an appropriate $(d+2)$ -dimensional index theorem. An outline as to how this may be done is given. In fact, the non-abelian anomaly can be related to the nontrivial topology of the orbit space [16, 17].

Suppose we consider again the action of a Yang-Mills theory with a chiral fermion on an even-dimensional Euclidean space compactified to S^{2p} . Let G be a simple, simply connected group such as $SU(n)$. The effective action $Z[A]$ is

$$e^{-Z[A]} = \int D\psi D\bar{\psi} e^{-\int d^{2p}x \bar{\psi} i \not{D} \psi}. \quad (40)$$

The identification of (40) with $\det(i / \not{D} P_+)$ is not possible since apart from the regularization problems, the operator $i / \not{D} P_+$ does not have a well-defined eigenvalue problem. This can be avoided by using the operator \not{D}^* with a Dirac spinor instead of a Weyl spinor. The new action allows us to express (40) as a determinant

$$e^{-Z[A]} = \int \mathcal{D}\psi \mathcal{D}\bar{\psi} e^{-\int d^{2p}x \bar{\psi} \not{D}^* \psi} = \det(i / \not{D}^*). \quad (41)$$

As we know, the functional integral is not gauge invariant

$$e^{-Z[A^g]} = e^{i\alpha_1[A,g]} - e^{-Z[A]}. \quad (42)$$

This lack of invariance has the effect that (41) cannot be expressed over the orbit space \mathcal{A}/\mathcal{G} .

The fermion determinant may itself be used to study the nontrivial topology of the orbit space. The variation has to be imaginary since its modulus is gauge invariant. Define now $/D_* = /D_+ + /D_- = /D + /A P_+$. It is the case that when a Weyl realization for the gamma matrices is used

$$\begin{aligned} (\det(i/D^*)) (\det(i/D^{*\dagger})) &= \det(i/D^* i/D^{*\dagger}) = \det[(i/D_+ + i/\partial_-)(i/D_- + i/\partial_+)] \\ &= \det(i/D_+ i/D_- + i/\partial_- i/\partial_+) = \det \begin{pmatrix} i/\partial_- i/\partial_+ & 0 \\ 0 & i/D_+/D_- \end{pmatrix} \end{aligned} \quad (43)$$

The factor $\det(i/\partial_- i/\partial_+)$ is ignored here, so it is found that $|\det(i/D^*)|^2$ is proportional to $\det(i/D_+ i/D_-)$. Since $\det(i/D_+ i/D_-)^2 = \det(i/D)^2$, where $/D$ is the usual Dirac operator, up to a constant, we obtain $|\det(iD^*)| = (\det(i/D)^{1/2})^{1/2}$. This can be regulated in a gauge-invariant manner. Thus, only the imaginary part of $\det(i/D^*)$ can change as implied by (42). As the real part always admits a gauge-invariant definition on general grounds, it can be said only the imaginary part may change. The real part has a gauge-invariant definition.

Let us try to look more deeply into the topological nature of the non-abelian gauge anomaly. It is important to examine the connection between the d -dimensional non-abelian anomaly and the $d+2$ index theorem which describes the abelian anomaly in $d+2$ dimensions. Let $g(\theta, x)$ be a family of gauge transformations that depend on a parameter $\theta \in S^1$ satisfying the boundary conditions $g(0, x) = g(2\pi, x) = e$. These define mappings $g : S^1 \times S^{2p} \rightarrow G$ classified by the homotopy classes $\pi_{2p+1}(G)$. Notice with these conditions on $g(\theta, x)$, the product S^1 and S^{2p} is topologically equivalent to S^{2p+1} . Let $A(x)$ be a gauge field such that it corresponds to a connection on a trivial bundle and let

$$A^\theta = g^{-1}(\theta, x)(d + A)g(\theta, x), \quad (44)$$

where $d = \sum_{\mu=0}^{d-1} dx^\mu (\partial/\partial x^\mu)$ is the resulting one-parameter family of group-transformed Yang-Mills configurations. Suppose the operator $iD_*(A)$ has no zero modes. Since just the phase can pick up an anomalous change under a gauge transformation, the operator $i/D^*(A)$ does not have them either for all $\theta \in U(1)$ as the modulus of the determinant is gauge invariant, so $|\det(i/D^*(A^\theta))| = |\det(i/D^*(A))|$, and we have

$$e^{-Z(A^\theta)} = \det(i/D^*(A^\theta)) = \det(iD^*(A))e^{i\alpha(A, \theta)}. \quad (45)$$

Of course, θ enters in α and in A^θ by means of $g(\theta, x)$, so the functional $\det(i/D_*(A^\theta))$ defines a particular mapping, $\theta \rightarrow e^{i\alpha[A, \theta]} \in U(1)$. These mappings can be characterized by means of an integral winding number,

$$k = \frac{1}{2\pi} \int_0^{2\pi} d\theta \frac{\partial \alpha[A, \theta]}{\partial \theta}. \quad (46)$$

The integrand is somewhat formal since it is not exact, otherwise $k = 0$. To actually get this winding number, it is necessary to extend A^g to a two-parameter family $A^{t,\theta}$ of gauge fields. If A^θ describes a circle S^1 in the space of gauge fields \mathcal{A} , then $A^{t,\theta}$ is defined on a two-dimensional disk W^2 with $\partial W^2 = S^1$

$$A^{t,\theta} = tA^\theta = tg^{-1}(\theta, x)(d + A)g(\theta, x), \quad (47)$$

with d as below (44), (t, θ) are polar coordinates of the disk W^2 , with $0 \leq t \leq 1$. On the boundary, $A^{t,\theta}$ becomes the one-parameter family of gauge-related configurations of (44). If A^θ describes a circle in \mathcal{A} , then $A^{t,\theta}$ describes a two-dimensional disk in the space of all gauge fields on S^{2p} that belongs to the trivial topological class in \mathcal{A} . This triviality is not important as far as the topology of the orbit space \mathcal{A}/\mathcal{C} . The determinant $\det(i/D^*(A^{t,\theta}))$ turns into a complex function of the gauge fields on the circle of gauge fields at its boundary. Omitting details of the calculation of the winding number, it may be shown it is equal to the difference between the positive and negative chirality modes of the ordinary Dirac operator i/D_{2p+2} in $d + 2$ dimensions

$$\text{ind}(i/D_{2p+2}) = n_+ - n_- = k = \int_0^{2\pi} d\theta \alpha[A, \theta], \quad (48)$$

where $i/D_{2p+2} = \sum_{\mu=0}^{2p+1} \gamma^\mu D_\mu$. The γ 's are now those of a $(d + 2)$ -dimensional space whose coordinates are t, θ, x^μ and $d\theta(\partial/\partial\theta) \equiv d_\theta$. This is executed by relating the zero modes $i/D^*(A^{t,\theta})$ to those of i/D_{2p+2} . In fact, the winding number (46) of the phase of the d -dimensional Weyl determinant is measured by the homotopy class in $\pi_{2p+1}(G)$ of mapping $g(\theta, x)$.

The next thing to do is evaluate $\text{ind}\left((i/D)_{2p+2}\right)$. Gauge field $A^{t,\theta}(x)$ is defined on the manifold $W^2 \times S^{2p}$ with a boundary $S^1 \times S^{2p}$, and so is the operator i/D_{2p+2} . To use the index theorem for manifolds without boundary, so no boundary corrections arise, consider the manifold $S^2 \times S^{2p}$. The Yang-Mills fields A on this larger manifold $S^2 \times S^{2p}$ are the pullbacks of connections on a principal bundle over $S^2 \times S^{2p}$ with a structure group G . So as far as the two sphere is concerned, we can use two local charts $D_\pm^2 \times S^{2p}$, where the disks D_+ and D_- are parametrized by (t, θ) , $((s, \theta))$ with W_+ being the previous W^2 . This should avoid singular parametrization. The region of overlap along the equator corresponds to $t = 1 = s$; the north (south) poles of S^2 are given by $t = 0, s = 0$. Two local gauge fields $(0, 0, A_\pm)$ on the two charts $W_+ (0 \leq t \leq 1)$ and $W_- (0 \leq s \leq 1)$ can be taken as

$$A_+(t, \theta, x) = t[g^{-1}(A + d + d_\theta)g] = A^{t,\theta} + tg^{-1}d_\theta g, \quad A_-(s, 0, x) = A. \quad (49)$$

The lower disk is trivial. Let us define an operator \bar{d} (by)

$$\bar{d} = d + d_\theta + d_t + d_s. \quad (50)$$

It can be seen that A_{\pm} at the equator $t = 1 = s$ are gauge related

$$A_+ = g^{-1}(\bar{d} + A_-)g \quad (51)$$

since $d_g(\theta, x) = 0 = d_g(\theta, x)$. Hence, A_{\pm} define a connection on the principal bundle over $S^2 \times S^{2p}$ with group G . So $g(\cdot, x) : S^1 \rightarrow \mathcal{C}$ is just the transition function between the local expressions A_{\pm} at the boundary $S^1 \times S^{2p}$ of the two patches. By the previous considerations, this also defines a mapping $g(\theta, x)$ homotopically equivalent to a mapping $g : S^{2p+1} \rightarrow G$. These transition functions define loops in \mathcal{C} that are classified by $\pi_1(\mathcal{C})$.

The expression of the index that needs to be computed for $p = d/2$ is taken to be

$$\text{ind}(i/D_{2p+2}(A)) = \int_{S^2 \times S^{2p}} \text{ch}_{p+1}(F) = \int_{W_+ \times S^{2p}} \text{ch}_{p+1}(F_+) + \int_{W_- \times S^{2p}} \text{ch}_{p+1}(F_-). \quad (52)$$

The Chern character $(2p+2)$ -form $\text{ch}_{p+1}(F)$ is closed and the local potential $(2p+1)$ -forms are given by the Chern-Simons forms Q^{2p+1} for the gauge fields on the corresponding charts. In fact, (51) is given by

$$Q^{(2p+1)}(A_+, F_+)|_{t=1} - Q^{(2p+1)}(A_-, F_-)|_{s=1}. \quad (53)$$

However, on account of (5.12), this expression is the variation of $Q^{(2p+1)}$ under a gauge transformation and given by $Q^{(2l-1)}(A^g, F^g) - Q^{(2l-1)}(A, F) = d\alpha^{2p-2}(A, F, a) + Q^{(2l-1)}(g^{-1}dg, 0)$ upon replacing l by $p+1$,

$$Q_+^{(2p+1)} - Q_-^{(2p+1)} = Q^{(2p+1)}(g^{-1}\bar{d}g, 0) + \bar{d}\alpha^{(2p)}, \quad (54)$$

where \bar{d} appears in (54) as the exterior derivative $\bar{d} = d + d_{\theta}$. The second term in this is an exact form and it does not contribute,

$$\text{ind}(i/D_{2p+2}) = (-1)^p \left(\frac{i}{2\pi}\right)^{(p+1)} \frac{p!}{(2p+1)!} \int_{S^1 \times S^{2p}} \text{Tr}(g^{-1}\bar{d}g)^{2p+1}. \quad (55)$$

In fact, this is the number of times $g(\theta, x)$ wraps around S^{2p+1} and is in $\pi_{2p+1}(G)$.

The local density $id_{\theta}\alpha[A, \theta]$, which describes the non-abelian anomaly in a d -dimensional spacetime can be identified and both sides of (48) have to be related. In terms of $Q^{(2p+1)}$, the index i/\mathcal{D}_{2p+1} reduces to

$$\text{ind}(i/\mathcal{D}_{2p+2}) = \int_{S^1 \times S^{2p}} Q^{(2p+1)}(A^{\theta} + \hat{v}, F^{\theta}), \quad (56)$$

where $\hat{v} \equiv g^{-1}d_{\theta}g = v d\theta$ and $F^{\theta} = \bar{d}(A^{\theta} + v)^2 = g^{-1}Fg$, since the term in Q evaluated at $s = 1$ does not have a $d\theta$ component and cannot contribute to the integral $S^1 \times S^{2p}$. The only term which contributes to (56) is the term linear in \hat{v} . By (48), the

local density for the anomaly is the local density for ind (i/\mathcal{D}_{2p+2}) . Thus, what is needed is the first-order variation

$$Q_1^{(2p)}(v, A^\theta, F^\theta) = Q^{(2p+1)}(A^\theta + v, F^\theta) - Q^{(2p+1)}(A^\theta, F^\theta). \quad (57)$$

As $Q^{(2p+1)}$ is known, this is not difficult to compute. For $d = 2p = 4$, using (48) this gives

$$\frac{1}{2\pi} d_\theta \alpha = \int_{S^4} Q_1^4(v, A^\theta, F^\theta) = -\frac{i}{48\pi^3} \int_{S^4} \text{Tr} \left(v d \left(A^\theta dA^\theta + \frac{1}{2} (A^\theta)^3 \right) \right). \quad (58)$$

Factoring $d\theta$ and setting $\theta = 0$, we find the expression for the anomaly

$$\frac{\partial \alpha}{\partial \theta} [A, g] = -\frac{i}{24\pi^2} \int_{S^4} \text{Tr} \left[v d \left(A dA + \frac{1}{2} A^3 \right) \right]. \quad (59)$$

From (29) and (31), it is found that

$$V_{(i)}[A](x) = D_\mu J_{(i)}^\mu = -\frac{i}{24\pi^2} \text{Tr} \left[T_{(i)} \epsilon^{\nu\rho\sigma\kappa} \partial_\nu \left(A_\rho \partial_\sigma A_\kappa + \frac{1}{2} A_\rho A_\sigma A_\kappa \right) \right]. \quad (60)$$

There are topological consequences related to these results. These considerations provide a check of the relation $\pi_1(\mathcal{C}) = \pi_5(G)$. What can be said about $\pi_2(\mathcal{A}/\mathcal{C})$. Recall the set of Yang-Mills potentials $A^{t,\theta}$ in \mathcal{A} . The potentials A^θ when $t = 1$ on the boundary are gauge related. The projection of this disk on the orbit space is a two-sphere since all A^θ are projected onto a reference potential A .

The t part of $A^{t,\theta}$ is topologically trivial so the two-sphere in \mathcal{A}/\mathcal{C} is always contractible $\pi_2(\mathcal{A}/\mathcal{C}) = 0$ if and only if the loop A^θ in \mathcal{A} is trivial so $\pi_1(\mathcal{C}) = 0$. The homotopic non-triviality of the gauge group is equivalent to that of the two-sphere. If not the homotopy used to contract the sphere could be used to deform the loop $g(\theta, x)$ to a point $g(x)$ and $\pi_1(\mathcal{C}) = \pi_2(\mathcal{A}/\mathcal{C})$. There is a global topological obstruction to removing the phase factor in the definition of the determinant when the non-triviality is present. This means the determinant cannot be restricted to the orbit space consistently, so there is no definition in a gauge-invariant manner. Instead, if $\det A$ is computed using a regularization procedure, it is found that

$$\det(A^g) = e^{i\alpha_1[A,g]} \det(A), \quad (61)$$

where α_1 can be considered a mapping of $\mathcal{A} \times \mathcal{C}$ in the multiplicative group of complex numbers, hence the one. The action of two transformations required that mod 2π ,

$$\alpha_1[A^g, g'] = \alpha_1[A, gg'] + \alpha_1[A, g] = 0. \quad (62)$$

This is a one-cycle condition. It relates topological properties with cohomology since the variation of the phase of the determinant is an obstruction to projecting gauge orbits in \mathcal{A} onto points in \mathcal{A}/\mathcal{C} . In this picture, a complex line bundle over \mathcal{A}/\mathcal{C} is defined, the determinant bundle, characterized by its first Chern class. This

determines an element in $\pi_1(\mathcal{C}) = \pi_{2p+1}(G) = \pi_2(\mathcal{A}/\mathcal{C}) = Z$. The determinant bundle over a non-contractible two-sphere in orbit space with a winding number in $\pi_1(\mathcal{C})$ is identical to the bundle describing a monopole of the same number of units of magnetic charge. Upon writing a representation of the first Chern class in a topologically nontrivial configuration is the same as giving a specific form of the anomaly.

In $d = 4$ the abelian anomaly is governed by $\pi_0(\mathcal{A}/\mathcal{C}) = \pi_3(G)$ and the non-abelian anomaly by $\pi_2(\mathcal{A}/\mathcal{C}) = \pi_5(G)$. In fact Witten's global anomaly depends on $\pi_1(\mathcal{A}/\mathcal{C}) = \pi_4(G)$ of course with $\pi_5(G) = Z(\pi_{2p+1}(G)) = Z$ is a sufficient but not necessary condition for the existence of an anomalous variation of Z [A]. The determinant may have a local variation even if the topology does not force it to acquire a nontrivial topological phase.

6. Conclusions

This work has given a practical way of extracting the local variation and with it the associated anomaly, which is not necessarily related to the existence of a nonzero integer winding number.


Hence, even if the variation of the phase functional does not have a global topological meaning for a compactified spacetime S^d , it still provides the physical perturbative expression for the non-abelian anomaly. It may appear as a topological obstruction in less topologically trivial classes in the local cohomology of gauge fields. This way of proceeding though has the consequence that it results in the cohomological descent procedure [8–21].

Author details

Paul Bracken
 Department of Mathematics, University of Texas, Edinburg, TX, USA

*Address all correspondence to: paul.bracken@utrgv.edu

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Atiyah MF. Topological quantum field theories. Publications Mathématiques de l'IHÉS. 1989;**68**:175-186
- [2] Atiyah MF, Hintchen NJ, Singer IM. Self-duality in four-dimensional Riemannian geometry. Proceedings of the Royal Society. 1978;**A 362**:425-461
- [3] Bailin DB, Love A. Introduction to Gauge Theories. Bristol: Institute of Physics Publications; 1993
- [4] Gribov VN. Quantization of non-abelian gauge theories. Nuclear Physics B. 1978;**159**:1-19
- [5] Bardeen WA, Zumino B. Consistent and covariant anomalies in gauge and gravitational theories. Nuclear Physics B. 1984;**244**:421-453
- [6] Chern SS, Simons J. Characteristic forms and geometric invariants. Annals of Mathematics. 1974;**99**:48-69
- [7] Coleman S. Aspects of Symmetry, Selected Erice Lectures. Cambridge: Cambridge University Press; 1985
- [8] Atiyah MF, Jones JDS. Topological aspects of Yang-Mills theory. Communications in Mathematical Physics. 1978;**61**:97-113
- [9] Treiman SB, Jackiw R, Zumino B, Witten E, editors. Current Algebra and Anomalies. Singapore: World Scientific; 1988
- [10] Witten E. Topological quantum field theory. Communications in Mathematical Physics. 1988;**117**:353-386
- [11] de Azcarraga JA, Izquierdo JM. Lie Groups, Lie Algebras, Cohomology and some Applications in Physics. Cambridge: Cambridge Monographs; 1998
- [12] Bertlmann RA. Anomalies in Quantum Field Theory. Oxford: Oxford Science; 2000
- [13] Fujikawa K. Path integral for gauge theories with fermions. Physical Review D. 1980;**21**:2848-2558
- [14] Bracken P. Calculation of anomalies in gauge theories using a geometric approach. International Journal of Modern Physics E. 2004;**13**:801-810
- [15] Bracken P. Quantization of anomalous Abelian gauge theories. Quantum Studies: Mathematics and Foundations. 2018;**5**:535-542
- [16] Pontryagin LS. Topological Groups. NY: Gordon and Breach; 1966
- [17] Nelson P, Alvarez-Gaumé L. Hamiltonian interpretation of anomalies. Communications in Mathematical Physics. 1985;**99**:103-114
- [18] Saravi RE, Muschietti F, Schaposnik FA, Solomin JE. Chiral symmetry and functional integral. Annals of Physics. 1984;**157**:360-393
- [19] Kopper C, Lévêque B. Regularized path integrals and anomalies: $U(1)$ chiral gauge theory. Journal of Mathematical Physics. 2012;**53**:022305
- [20] Varshovi A. Geometry and topology of anti BRST symmetry in quantized YM gauge theories. International Journal of Geometric Methods in Modern Physics. 2022;**19**:2250007
- [21] Percacci R. On the quantization of a theory with global gauge anomalies. Modern Physics Letters. 1987;**A 2, 12**: 977-982

Topological Phenomena in Spin Systems: Textures and Waves

Paula Mellado and Roberto E. Troncoso

Abstract

This chapter reviews the implications of topology in the static and dynamics of magnetic systems. Our focus is twofold. In the first part, we describe how the application of topology allows an understanding of the structure and dynamics of magnetic textures that separate different magnetic domains in magnetic materials. Topological textures are rationalized in terms of elementary topological defects that determine complex magnetic orders and magnetization dynamics processes in the underlying magnetic systems. The second part studies topological phases and topological phenomena associated with the band theory of linear magnetic excitations. Topological spin waves are usually accompanied by exotic phenomena in magnetic materials such as the emergence of chiral edge states and the magnon Hall effect.

Keywords: topological defects, domain walls, skyrmions, topological magnons, winding number, edge states, hall effect, chiral modes

1. Introduction

Topology in condensed matter physics categorizes the robustness of classical or quantum states under smooth deformations. Dissimilar physical states may manifest such robustness in terms of a common set of preserved quantities. The preserved quantities, or invariants, affect the static properties of such systems and their dynamics and allow the classification of their physical states. Crucially, the study of topological invariants in such systems allows the definition of robust rules that facilitate the understanding and modeling of the associated physical phenomena.

The application of topology to condensed matter physics began with the study of lattice defects like dislocations and disclinations. In magnetism, the topological methodology allowed the systematic study of complex magnetic structures that take the form of localized magnetic defects in real space. In a medium with an underlying magnetic order, a defect is a point or a collection of points where the order parameter, usually a function of the magnetization vector, is not well defined. The defect is topological when it cannot be removed, whatever the continuous modifications exerted on the order parameter distribution over space. Depending on the system's geometry, the external fields, and the interactions between its constituents, topological spin textures may arise aimed to separate distinct magnetic phases in a sample. Examples of such textures are domain walls (DWs). In other cases, the spin textures

can be part of the magnetic order and define the magnetic landscape of the whole system. An example could be a lattice of periodically arranged skyrmions, or a lattice that relaxes into chiral spin textures periodically arranged. In either case, the study of self-localized topological spin textures and the rules associated with their creation, merging, and motion, strongly rooted in their winding numbers, has been extremely useful in characterizing the ground state and dynamics of magnetic systems.

Research related to the transport of topological textures in real space has set the basis for spintronics and racetracks memory devices. The first part of this chapter is devoted to reviewing the main results regarding topological textures in real space. Though they may arise in ferromagnets, antiferromagnets, or ferrimagnetic systems, here, we focus on topological spin textures on ferromagnetic samples, which constitute a mature field of research.

More recently, topology has caused renewed interest because of the study of topological phase transitions and topological phases of matter [1]. These phases usually manifest in reciprocal space, where the nontrivial wrapping of the Brillouin zone around the Hamiltonian space leads to exotic topological states. In this context, topological magnetic excitations in the form of spin waves or magnons (the quanta of spin waves) have recently become a very active and broad field. Transport with spin waves can be accomplished without electronic degrees of freedom, reducing the Joule heating, and making the spin a promising route to engineer the transmission of information. When spin waves propagate unidirectionally along the surface of a system, they are topologically protected by the bandgap in the reciprocal space. These chiral spin waves very often imply robustness because they can only be destroyed through external forces that consume a finite amount of energy. Topological bands are usually associated with exotic transport phenomena like the thermal Hall effect. The study of topological spin excitations has led to a new field called topological materials, including topological insulators and topological semimetals.

We begin by briefly summarizing the main interactions among spins in magnetic systems. The hierarchy of such interactions together with the geometrical properties of the underlying magnetic system gives rise to anisotropies that constrain the spin vector along specific directions.

1.1 Magnetic interactions

The different coupling mechanisms between spins can give rise to various ground states of spin textures and dynamics. In the following, we consider a semiclassical model of spins represented by vectors localized at lattice sites. In the so-called micromagnetic limit, the system of localized spins is transformed into a smoothly varying spin texture, described by a set of continuous vector fields. Next, we describe the most relevant magnetic interactions involved in the stabilization of topological defects and the realization of topological bands in magnetic materials. These include the exchange, Dzyaloshinskii-Moriya (DM), magnetostatic (dipolar), and Zeeman interactions.

1.1.1 Exchange interaction

It originates from the concomitance of the Pauli exclusion principle and the Coulomb interactions. In the classical Heisenberg model, where the exchange interaction is the only term, the Hamiltonian is generally written as

$$H_E = \sum_{\alpha\beta=x,y,z} \sum_{\langle \mathbf{r}\mathbf{r}' \rangle} J_{\mathbf{r}\mathbf{r}'}^{\alpha\beta} S_r^\alpha S_{r'}^\beta, \quad (1)$$

with $J_{\mathbf{r}\mathbf{r}'}$ the exchange coupling constant among spins S_r and $S_{r'}$ localized at positions \mathbf{r} and \mathbf{r}' . Here, $\langle \mathbf{r}\mathbf{r}' \rangle$ denotes nearest-neighbor (NN) spins, revealing the interaction's short-ranged nature. For materials that are isotropic, the spatial and spin dependence of the exchange tensor can be suppressed, and the exchange coupling can be treated as a homogeneous coupling constant between spins at neighboring sites. The sign of J determines, among other ingredients, the emergence of ferromagnetism ($J < 0$) or antiferromagnetism ($J > 0$).

1.1.2 Dzyaloshinskii-Moriya interaction

The Dzyaloshinskii-Moriya (DM) interaction is an asymmetric exchange coupling that favors a noncollinear orientation of spins. It was proposed to be the source of a nonzero spontaneous magnetic moment in certain magnetic states of the magnetic material Hermitite, by Dzyaloshinskii in 1958 [2]. The microscopic origin of the effect was later found by Moriya [3] who showed how spin-orbit coupling gives rise to an antisymmetric interaction mechanism in systems with low magnetic order. The DM spin Hamiltonian reads

$$H_{DM} = \sum_{\langle \mathbf{r}\mathbf{r}' \rangle} \mathbf{D}_{\mathbf{r}\mathbf{r}'} \cdot (\mathbf{S}_r \times \mathbf{S}_{r'}), \quad (2)$$

with $\mathbf{D}_{\mathbf{r}\mathbf{r}'}$ the DM vector coupling. The direction and magnitude of $\mathbf{D}_{\mathbf{r}\mathbf{r}'}$ are determined by the symmetry class the crystalline lattice belongs to. The DM interaction favors a canting among neighbor spins that competes with the direct exchange interaction. Minimizing the magnetic energy (containing DM) leads to twisted magnetic structures, such as skyrmions.

1.1.3 Magnetostatic interaction

The magnetostatic energy is also called dipolar interaction energy. In a crystal, each magnetic moment creates a dipolar field, and each moment is exposed to the magnetic field created by all other dipoles. Denoting by \mathbf{H}_r the magnetic field created by all spins other than spin r in the position of \mathbf{S}_r , the magnetostatic energy can be written as follows:

$$E_M = -\frac{\mu_0}{2} \sum_r \mathbf{S}_r \cdot \mathbf{H}_r, \quad (3)$$

where μ_0 is the magnetic permeability. In a continuum model, the magnetostatic energy corresponds to the long-ranged interaction energy, equal to the work made against the magnetic field generated by a continuous magnetic moments distribution, to bring an elementary magnetic moment from infinity to its actual position. The magnetostatic field at a given location within the body depends on the contributions from the whole magnetization vector field. The magnetostatic energy is a nonlocal interaction and can be taken into account by introducing the appropriate magnetostatic field according to Maxwell equations for magnetized media. In a discrete model,

the dipolar interactions between spins are computed by considering spins as point dipoles located at the sites of a lattice [4].

1.1.4 Zeeman interaction

It is the interaction energy of the magnetic moment of a system with the external field \mathbf{H}_{ext} . It is defined as

$$E_z = -\mu_0 \sum_r \mathbf{S}_r \cdot \mathbf{H}_{ext}. \quad (4)$$

2. Static of magnetic textures

2.1 Topological defects

Ordered phases of matter, such as magnetism or superconductivity in solid state systems, are characterized by a *parameter space* V . This is the space where the order parameter lives in and shapes the spatial and temporal features of the phase. In magnetism, $V = \mathbb{S}^{n-1}$, n is the number of magnetization components, and a single spin in three dimensions (3D) is represented by a vector \mathbf{S} . For $n = 3$, the parameter space is the 2-sphere (two dimensional) \mathbb{S}^2 . A m -sphere is the space of all elements, x , in $(m + 1)$ D satisfying $|x| = 1$ [5].

A *topological defect* represents regions where the order parameter is not well defined or regions where it describes more complex structures in space. These objects can be explored by mathematical tools from *homotopy theory* [6, 7]. For every parameter space V , the r th homotopy group $\pi_r(V)$ exists, and each group element is a homotopy class. Topologically different spaces have different classes of equivalent objects. Members from different topological classes cannot be deformed into another [5].

One of the simplest homotopy groups is $\pi_1(\mathbb{S}^1)$, the group of mappings from the 1-sphere (a circle) to the 1-sphere. $\pi_1(\mathbb{S}^1)$ is called the fundamental group and is isomorphic to \mathbb{Z} , which is the fundamental group of the circle. This means that we assign the number of times it goes around the circle to any loop [8]. Physically, it indicates all the possible ways to wrap a circle with a loop ($r = 1$). A simple class is the trivial one that corresponds to the possibility of shrinking the loop to a point. The existence of nontrivial classes in a particular homotopy group indicates the presence of topological defects in the physical systems that realize it. Vortices, magnetic textures in two or three dimensions where the order parameter is null at the core, are good examples of the realization of $\pi_1(\mathbb{S}^1)$.

In higher dimensions, the space of parameters features $V = \mathbb{S}^m$, and the homotopy group becomes $\pi_n(\mathbb{S}^m)$. The most notorious example in magnetism occurs in two dimensions and corresponds to skyrmion textures [9, 10]. Here, the parameter space is a 2-sphere, and two angles determine the coordinate space. Together they configure the second homotopy group $\pi_2(\mathbb{S}^2)$ which is, again, isomorphic to \mathbb{Z} . In other words, the order parameter \mathbf{S} takes values from the coordinate space, parameterized by the spherical angles θ and ϕ , that is, \mathbb{S}^2 , and sends them to parameter space \mathbb{S}^2 . The winding number is the number of times the order parameter wraps \mathbb{S}^2 . Extensions of it include higher-dimensional skyrmions [11], monopoles [12], and Hopfions [13].

Generally, for Ising, XY, and Heisenberg magnets, the nontrivial homotopy groups are the following. Ising: $\pi_0(\mathbb{S}^0) = \mathbb{Z}_2 \equiv 0, 1$. In the case of Ising spins, topological defective surface exists in 3D, topological line defects in two dimensions, and topological point defects in 1D. For XY spins $\pi_1(\mathbb{S}^1) = \mathbb{Z} \equiv 0, +1, -1, +2, -2, \dots$ there exist topological linear defects in 3D, and topological point defects in 2D, examples of topological point defects in two dimensions are the Kosterlitz-Thouless vortices and antivortices [14]. For Heisenberg spins, $\pi_2(\mathbb{S}^2) = \mathbb{Z}$. This includes topological defects in the form of points in 3D, as the Bloch point [15]. The Bloch point and the vertical Bloch line are elementary topological defects in topological textures like domain walls. In the Bloch point, every magnetization orientation appears once exactly. The vertical Bloch line is a topological soliton and separates two different magnetization directions inside domain walls. The singular vortex, a topological point defect for XY spins, consists of a cut through the length of a Bloch point [15].

Magnetic solitons, nonlinear excitations in magnetic systems, are shape-preserving and self-localized magnetic structures. Typical examples include the magnetic vortex, bubble, skyrmion, and domain wall [12]. They have a small size and high stability and can be reconstructed from elementary topological defects. The topological constraints from the associated homotopy group directly influence the properties of materials by catalyzing or inhibiting the switching between different magnetic-ordered states by means of topological defects present in magnetic solitons [16]. Consider the homotopy group $\pi_2(\mathbb{S}^2)$. For Heisenberg spins, it considers a topological point defect in a three-dimensional medium and a topological soliton for a medium in 2D. The transformation from one topological soliton into another occurs through the exchange of topological defects with topological indices equal to the difference between the indices of the two solitons.

In the following, we will focus on the origin, stabilization, and dynamics of domain walls and two-dimensional skyrmions.

2.2 Domain walls

Domain walls (DWs) are transition regions between different domains. A domain wall arises when the boundary conditions are not uniform. This is the case of finite samples, where boundary conditions are imposed by the energetically preferred magnetization orientations at the edges [17], like in disks and nanostrips.

Whether or not a DW can be continuously transformed one into the other depends on the nature of the first homotopy group of the order parameter space $\pi_1(V)$: if it is trivial, then all walls are topologically equivalent, but if it is not, then topologically different walls are allowed to exist. The only nontrivial case occurs for the XY spins, $\pi_1(\mathbb{S}^1)$.

2.2.1 Domain walls in wires

For XY spins, there is an infinite number of topologically different DWs. This is not the case for Heisenberg spins where the magnetization vector is allowed to tilt out of the plane and each DW path can continuously deform into another within the unit sphere surface of the magnetization vector \mathbf{M} . An example of DW made out of XY spins in a wire is the 180°-DW, which separates two adjacent oppositely oriented domains. In the case of domains with collinear spins, they can be of the head-to-head or tail-to-tail types [18]. For domains with parallel moments, there are the up-down,

down-up Neel DW and the up/down, down/up Bloch DW [19]. Using a polar angle to an axis orthogonal to the wire, a topological index or winding number (or topological charge) for these DW can be defined

$$W = \frac{1}{2\pi} \int_{-\infty}^{\infty} \frac{d\theta}{dz} dz = \frac{\theta(\infty) - \theta(-\infty)}{2\pi}. \quad (5)$$

W describes how many times \vec{m} wraps the unit circle when the spatial coordinate sweeps the magnetic wire. For head-to-head (tail to tail) walls, $W = 1/2$ ($W = -1/2$). In the presence of an external field along the wire, a DW with $W = 1/2$ ($W = -1/2$) moves along (against) the field to lower the Zeeman energy. When present at the edges of magnetic films, these domain walls, combined with other topological defects, may give rise to composite domain walls as the transverse and vortex domain walls. For Heisenberg spins in a wire, all DWs with the same boundaries have the same winding number.

W of a single domain is zero, and thus, it is defined as topologically trivial. The 1D model is a good approximation for small-diameter (much less than exchange length and the DW width) nanowires or large bulky magnets whose magnetization varies only along one direction.

2.3 Elementary defects in flat magnets

In nanomagnets, elementary defects consist of vortices with integer winding numbers and edge defects with half-integer winding numbers. In this framework, domain walls are composite objects made out of two or more elementary defects [20].

For two-dimensional XY finite magnets, the bulk winding number along a path C in \mathbb{R}^2 is $W = \frac{1}{2\pi} \int_C \nabla \theta \cdot dr$. If $W = 0$, there is no singularity inside C . If $W = \pm 1$, there is a singularity inside C called vortex (antivortex). At the edge of the strip, C can be a segment of such edge, and the winding number is modified to $W = \frac{1}{2\pi} \int_C \nabla(\theta - \theta_c) \cdot dr$, where θ_c is the angle of the sample edge. W takes fractional values of $\pm 1/2$ in edge defects. The sum of all winding numbers over the bulk and edge of the sample is conserved. The total winding number of bulk and edges is a topological invariant [21].

In a ferromagnet without intrinsic anisotropy, the magnetic energy consists of the exchange contribution and the magnetostatic energy. The magnetic field and magnetization vector are related through Maxwell equations. Analytical treatment aimed to find magnetic configurations of the lowest energy is possible in a thin film limit where magnetization lies in the plane of the film. In this case, the magnetic energy becomes a local functional of the magnetization composed of two terms: the exchange term, which corresponds to the XY model with the ground states obeying the Laplace equation $\nabla^2 \theta = 0$ in bulk, and the term that expresses the magnetostatic energy due to magnetic charges at the film edge and sets the boundary conditions [22]. In the limit when the width w of the sample (with thickness t and exchange length λ) is larger than the effective magnetic length $\Lambda = 4\pi\lambda^2/t \log(w/t)$, magnetization at the edge is forced to be parallel to the boundary, in the thin film geometry ($t \ll w$) [20].

Solutions in this limit have been constructed by analogy with electrostatic in two dimensions [20]. Angle gradients are associated with components of the electric field, and topological defects with winding numbers $+1$ and -1 become positive and negative point charges with unit strength. In the presence of an edge, there are two classes of solutions: (1) single vortices, which are repelled from the boundary, and (2) half

vortices, also called edge defects, with a singularity at the edge and winding number $\pm 1/2$. A superposition of two edge defects with winding number $-1/2$ and one vortex with charge $+1$ has zero total topological charge [20]. The conservation of a topological charge constrains the creation and annihilation of DW.

2.4 Composite domain walls in soft magnetic strips and rings

Domain walls in magnetic strips and rings are composite objects made of two or more elementary defects: vortices with integer winding numbers ($W = \pm 1$) and edge defects with fractional winding numbers ($W = \pm 1/2$) [22]. This framework provides a basic understanding of the complex switching processes observed in ferromagnetic nanoparticles [21]. In ferromagnets, the competition between exchange and magnetic dipolar energies creates nonuniform magnetization patterns in the ground state: whereas the exchange energy favors a state with uniform magnetization, magnetic dipolar interactions align the magnetization vector with the surface. In a large magnet, a compromise is reached by forming uniformly magnetized domains separated by domain walls.

2.4.1 Transverse DW

In a strip of width w and thickness t , a domain wall interpolates between the $\theta = 0$ and $\theta = \pi$ ground states and can be constructed out of two edge defects with opposite winding numbers. Its solution is $\tan(\theta(x, y)) = \pm \frac{\cos(\pi y/w)}{\sinh(\pi(x-X)/w)}$ [22]. From the electrostatic analogy, the two defects experience an attractive Coulomb force (not strong enough to overcome edge confinement). This force holds the composite domain wall together [21] (see **Figure 1**). In wider and thicker strips ($wt > \lambda^2$), the magnetostatic energy breaks the symmetry between the defects with positive and negative winding numbers. In particular, the $+1/2$ edge defect has higher magnetostatic energy than their $-1/2$ counterparts [23].

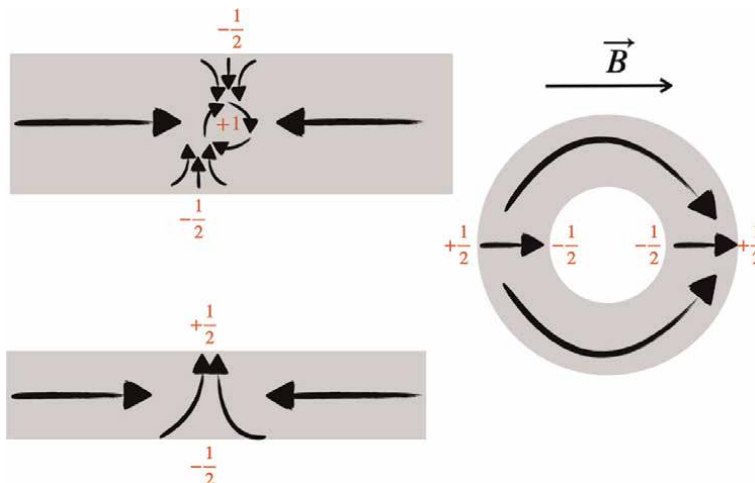


Figure 1. Vortex DW consisting of two edge defects and one vortex defect. Transverse DW consisting of two edge defects with the same winding number. Narrow nanoring after the applied field \vec{B} has created two domain walls made out of two edge defects each.

2.4.2 Vortex DW

Vortex walls are stabilized when a strip's width and thickness substantially exceed the exchange length [24]. In this limit, the magnetostatic energy is the dominant contribution to the energy of a domain wall and the primary force determining the shape of topological defects. In the geometry of a thin film, the magnetostatic energy is minimized if the density of magnetic charges vanishes in bulk, and if on the surface the magnetization follows the contour. Domain walls in this limit are made of two $-1/2$ edge defects and a $+1$ vortex between them. Although different in shape, the $-1/2$ edge defects in the exchange and magnetostatic limits have identical topological properties (see **Figure 1**).

Transverse and vortex walls have the same total winding number equal to zero. Thus, the transverse wall and the vortex wall are topologically equivalent.

Conservation of topological charge has important implications for magnetization dynamics in nanomagnets. When a magnetic field is more significant than a specific value, but well below the Walker breakdown [25], a series of DW structure transitions occur to preserve the total winding number of the DW. For instance, edge defects with $-1/2$ winding number can only give birth to an antivortex of winding number -1 after changing its own winding number to $+1/2$ [26]. In rings, the ground state contains no topological defects and has zero magnetic dipole moment. The ring can be magnetized by applying an in-plane magnetic field. Switching off the field leaves the ring in a metastable state with remnant magnetization containing two composite domain walls [26]. By applying the magnetic field in the opposite direction, the walls can be set in motion on a collision course and may annihilate, leaving the magnet in the ground state. However, direct annihilation of two domain walls is impossible since both have edge defects with winding number $-1/2$ at the inner edge of the ring. Therefore, in thin and narrow rings where vortices are forbidden energetically, the two domain walls do not annihilate but instead form a 360°-DW (see **Figure 1**). In thicker and wider rings, annihilation does occur, and the ring returns to a ground state. In this case, since the edge defects cannot migrate into the bulk, they alter their signs by exchanging a vortex. The $+1/2$ defect emits a $+1$ vortex into the bulk and converts into a $-1/2$ defect [27].

2.5 Defects in disks with XY and Heisenberg spins

In soft magnetic films in two dimensions, the magnetostatic energy dominates the anisotropy terms, so the magnetization is confined to the plane of the film. Perfect confinement means that spins are XY. As confinement is not perfect in actual samples (with finite magnetization and nonzero exchange), topological defects as the vortex with topological charge $W = +1$ and the antivortex with $W = -1$ are regularized by having the core magnetization in the third dimension, over a size proportional to the micromagnetic exchange length $\Lambda = \sqrt{2A/(\mu_0 M_s^2)}$, where A is the exchange constant and M_s is the saturation magnetization. Owing to the nonzero winding number, an isolated vortex cannot be a localized object in an infinite film. However, they can exist locally in confined geometries such as nanodisks. Indeed, these structures are topologically stable if the magnetization is assumed to stay in the plane at infinity. This is the case of a disk with magnetization at the edge in the plane. Physically, as the edge constraint arises from magnetostatics, the avoidance of surface magnetic charges dictates that \vec{m} be tangent to the edge. Therefore, for a disk-

shape sample, the stabilized winding number is 1. For Heisenberg spins inside a disk, the core of the vortex is regular with purely out-of-plane moment $m_z \sim p = \pm 1$, with p the polarity of the vortex. Because of the topology of the edge magnetization, the vortex cannot be continuously erased into a uniform structure, and the polarity cannot continuously change from +1 to -1. Thus, the constraints at the magnetic film's edge let other topological structures appear. The Heisenberg vortex or antivortex is also known as meron [7].

2.6 Skyrmion textures

Skyrmions, swirling spin textures appearing in chiral magnets, are magnetic structures in which the spins point in all the directions wrapping a sphere [9, 28]. These topological spin textures have ignited a growing interest in spintronics [10] due to their rich phenomenology as well as novel potential applications [29]. Their nanoscale size, topologically protected stability [28], and the very low electric current densities needed to displace them [30] are among the best qualities that make them attractive candidates for information carriers in high-density data-storage technologies. They have been explored in many magnetic materials, also named *chiral magnets*, for example, MnSi [30–32], Fe_{1-x}Co_xSi [33–35], Mn_{1-x}Fe_xGe [36], FeGe [37], La_{0.5}Ba_{0.5}MnO₃ [38], and CuOSeO₃ [39], among others [12]. Spontaneous skyrmion phases have been synthesized using temperature and external magnetic fields. Their detection is typically carried out by neutron scattering [31], Lorentz transmission electron microscopy (LTEM) [37], and spin-resolved scanning tunneling microscopy [40] experiments.

The spin texture of magnetic skyrmions can originate from various mechanisms. Depending on the particular magnetic system, skyrmions can be stabilized by long-ranged dipolar, Dzyaloshinskii-Moriya, frustrated exchange, or four-spin exchange interactions [10]. In chiral magnets, spin orbit coupling and lack of inversion symmetry cause the appearance of the Dzyaloshinskii-Moriya interaction [2, 3]. This interaction is responsible for the stability of static skyrmions in two and three dimensions [28]. In magnetic systems, single skyrmions as well as the skyrmion crystal can become the lowest energy configuration by adjusting temperature and the applied magnetic field.

A (single) skyrmion configuration is a texture described by the direction of spin $S(\mathbf{r})$ at spatial position $\mathbf{r} = (x, y)$, where the spin at the core points down, while at the perimeter point up, as shown in **Figure 2**. The topological skyrmion number is defined as a measure of the wrapping of $S(\mathbf{r})$ around a unit sphere [5]

$$Q = -\frac{1}{4\pi} \int \mathbf{m} \cdot \left(\frac{\partial \mathbf{m}}{\partial x} \times \frac{\partial \mathbf{m}}{\partial y} \right) d^2 \mathbf{r} \quad (6)$$

where 4π accounts for the surface of the sphere and \mathbf{m} is the local normal to the sphere, both partial derivatives of \mathbf{m} belong to the local tangent plane of the sphere and the modulus of their vector product is the sine of their angle. In Eq. (6), the integrand is the solid angle spanned by \mathbf{m} when one moves all spins to the Bloch sphere center in the spin space. In the case of cylindrical symmetry, with the spherical angles relative to the magnetization at infinity being sole functions of the radius r and the in plane angle, a useful relation is obtained between Q , p (the polarity of the core of the structure) and $S = \frac{1}{2\pi} \int_0^{2\pi} \frac{d\Phi}{d\phi} d\phi$, the winding number of the planar magnetization component: $Q = Sp$.

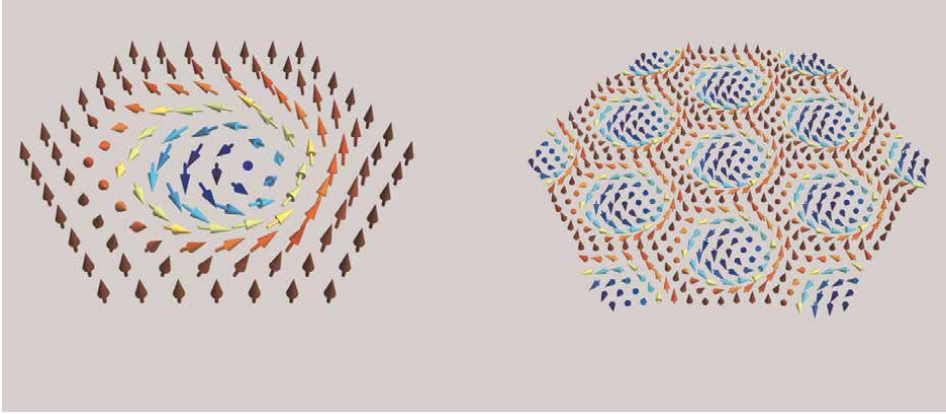


Figure 2. Schematic profile for single (left) and lattice (right) of magnetic skyrmions. The color indicates the relative direction of each spin respect to the vertical axis. Spins in dark red are pointing outward and those in blue are directed inward.

Topological protection does not mean absolute stability. The energy barrier between a skyrmion state and the single domain state is finite. Skyrmions can be generated from the edges of a sample without breaking the conservation of topology and from the bulk with the broken conservation of topology [41].

Up to now, we have discussed the statics of magnetic textures. We show next that the dynamics of magnetic textures can be affected by topology. Magnetization dynamics in continuous media is governed by the Landau-Lifshitz-Gilbert (LLG) equation supplemented by additional torque terms in the presence of spin-polarized currents. In this equation, the energy density gives rise to an effective magnetic field defined by a variational equation. This field originates from all the energetic contributions described above. Consequently, in most cases, the Landau-Lifshitz-Gilbert equation is not solvable analytically. Thiele's equation gives an approximate analytical solution for the dynamics of magnetic textures. To illustrate these ideas, we consider the dynamics of skyrmions next.

2.7 Dynamics and fluctuations of skyrmion textures

Skyrmions can be driven by external forces derived from charge currents, thermal gradients, and magnetic fields, among others. Particularly, current-driven skyrmion dynamics displays interesting topological transport properties, such as the *Skyrmion Hall effect* [32]. This effect results from the so-called spin-transfer torques phenomena, which is a torque exerted by carrier spins on the magnetization [42]. Driven skyrmion dynamics can also be triggered by mechanisms such as thermal gradients [43], inhomogeneity in the fields [29, 44], or magnon currents [45].

A first approximation captures the evolution of skyrmions in terms of a reduced set of collective coordinates. This approach is general and helps to parameterize the magnetization of the texture in terms of few degrees of freedom. Through the Landau-Lifshitz-Gilbert (LLG) equation, the skyrmion dynamics is mapped to a particle-like equation of motion, also known as Thiele's eq. [46]. Next, we outline this approach.

The starting point is the Landau-Lifshitz-Gilbert (LLG) eq. [47–49] for the magnetization direction \mathbf{M} , which incorporates spin-transfer torques [42, 50]. This torque

splits into adiabatic and nonadiabatic [42, 51–53] contributions, defined as $-\mathbf{v}_s \cdot \nabla \mathbf{M}$ and $\beta \mathbf{v}_s \cdot \nabla \mathbf{M}$, respectively. Here, $\mathbf{v}_s = -(pa^3/2eM)\mathbf{j}$ is the spin velocity of the conduction electrons, p is the spin polarization of the electric current, $e(>0)$ the elementary charge, and β a dimensionless parameter that accounts for the nonadiabaticity of the spin transfer.

The LLG equation reads

$$\left(\frac{\partial}{\partial t} + \mathbf{v}_s \cdot \nabla\right) \mathbf{M} = \mathbf{M} \times \mathbf{H}_{\text{eff}} + \alpha \mathbf{M} \times \left(\frac{\partial}{\partial t} + \frac{\beta}{\alpha} \mathbf{v}_s \cdot \nabla\right) \mathbf{M}, \quad (7)$$

where \mathbf{H}_{eff} is the effective field and α is the Gilbert damping constant. To describe the dynamics of single skyrmions, a particle-like motion [54–56] is considered (this approach is also useful for the case of domain walls [24]). It consists of a parametrization of the magnetization vector \mathbf{M} in terms of collective coordinates. For a single skyrmion moving rigidly along the trajectory $\mathbf{x}(t)$, we take as an ansatz the profile $\mathbf{M}(\mathbf{r}, t) = \mathbf{M}_0(\mathbf{r} - \mathbf{x}(t))$, where \mathbf{M}_0 represents the static skyrmion texture centered at the origin. The skyrmion profile \mathbf{M}_0 is obtained by minimizing the magnetic energy. Plugging in the ansatz on the LLG equation and integrating over complete space, we get for the skyrmion dynamics the Thiele's equation

$$a_{ij} \dot{x}^j(t) + F_i(\mathbf{x}(t)) = 0, \quad (8)$$

with the 2×2 -matrix $a_{ij} = -\varepsilon_{ikj} g_k + \alpha D_{ij}$. Here, indices i, j , and k denote coordinates x, y , and z , respectively, ε_{ijk} is the Levi-Civita symbol, and summation over repeated indices is assumed. The first term in a_{ij} contains the gyromagnetic vector defined by

$$g_i = -\frac{1}{2} \varepsilon_{ijk} G_{jk}, \quad (9)$$

where $G_{ij} = \int d\mathbf{r} \varepsilon_{klm} M_0^k \partial_i M_0^l \partial_j M_0^m$. This term in the equation of motion describes the Magnus force [30] exerted by flowing electrons. For a single skyrmion $g_i = g \delta_{iz} = 4\pi Q$, where Q is the skyrmion charge, that for our case is $Q = -1$. On the other hand, the second contribution represents the dissipative force whose components are $D_{ij} = \int d\mathbf{r} \partial_i M_0^k \partial_j M_0^k$, which obeys for the single-skyrmion case $D_{ij} = D \delta_{ij}$ because of the symmetry of the spin configuration.

The drift velocity of the skyrmion is affected, on one side, by a force \mathbf{F} given by $F_i(\mathbf{x}) = [\varepsilon_{ijk} g_j - \beta D \delta_{ik}] v_s^k - \partial V[\mathbf{x}] / \partial x_i$, that is explicitly written as

$$F_x = (-g v_s^y - \beta D v_s^x) - \partial V / \partial x, \quad (10)$$

$$F_y = (g v_s^x - \beta D v_s^y) - \partial V / \partial y, \quad (11)$$

containing both the gyrotropic and dissipative contribution due to the electron's coupling and a force due to the potential $V[\mathbf{x}]$ from the surrounding environment, for example, magnetic impurities, local anisotropies, or geometric defects. Potential $V(\mathbf{x}) = V_H(\mathbf{x}) + V_A(\mathbf{x})$ derives from an inhomogeneous magnetic field $\mathbf{H}(\mathbf{x})$ coupled to the magnetization of the ferromagnet and a position-dependent perpendicular anisotropy $A(\mathbf{x})$, with $V_H(\mathbf{x}) = -\frac{1}{\hbar} \int d\mathbf{r} \mathbf{M}_k(\mathbf{r} - \mathbf{x}) H_k(\mathbf{r})$ and

$V_A(\mathbf{x}) = -\frac{1}{\hbar} \int d\mathbf{r} M_z^2(\mathbf{r} - \mathbf{x}) A(\mathbf{r})$, respectively. The motion of single skyrmions is, therefore, determined by the solutions of Thiele's equation, with the forces given by Eqs. (10) and (11). Generalizations are straightforward and include random dynamics due to thermal fluctuations or disorder and noncoherent dynamics, that is, the motion with deformable shapes. The last leads to inertial terms in Thiele's equations, with an effective mass quantifying the degree of deformation.

We conclude this section by remarking that skyrmions have better stability than vortices. However, as we have seen, their nontrivial topological number leads to a finite gyrovector in the Thiele equation and, consequently, to the existence of a transverse motion in longitudinal driving force. Thus, there exists a threshold current density above which skyrmions can annihilate at the film edge. This edge effect strongly limits the speed of skyrmion propagation, which is vital for real applications.

2.8 3D skyrmions and beyond

Generalizations of 2D skyrmions are diverse. Particular efforts have been directed to realize three-dimensional textures in magnetic materials [12]. As an example we mention the magnetic hopfion [13, 57] and the three-dimensional skyrmion [11], a class of textures classified according to homotopy groups $\pi_2(\mathbb{S}^3)$ and $\pi_3(\mathbb{S}^3)$. These topological textures are characterized by the Hopf charge Q_H , given by

$$Q_H = \int d\mathbf{r} \mathbf{A}(\mathbf{r}) \cdot \mathbf{B}(\mathbf{r}), \quad (12)$$

where $\mathbf{B}(\mathbf{r}) = \varepsilon_{ijk} \mathbf{S} \cdot (\partial_j \mathbf{S} \times \partial_k \mathbf{S})$ is the emergent magnetic field and $\mathbf{A}(\mathbf{r})$ is the associated vector potential $\mathbf{B} = \nabla \times \mathbf{A}$. And the skyrmion number Q_s

$$Q_s = \frac{1}{24\pi^2} \int d\mathbf{r} \varepsilon^{\mu\nu\lambda} \text{tr}(\mathcal{L}_\mu \mathcal{L}_\nu \mathcal{L}_\lambda), \quad (13)$$

where $\mathcal{L}_\mu = \mathcal{R}^{-1} \partial_\mu \mathcal{R}$ and \mathcal{R} is a rotation matrix that locally aligns each spin along the quantized axis. These spin textures represent sophisticated structures even in their simplest form $Q_H = Q_s = 1$. Hopfions and 3D skyrmions lead to interesting dynamical properties such as the current-induced dynamics [13], field-driven resonance modes [58], and thermally activated drag of spin current [11]. 3D structures provide a versatile platform to explore the role of emergent phenomena. However, questions regarding the stability, nucleation, and low-energy dynamics are still under scrutiny and deserve special attention.

3. Topology in the dynamics of linear magnetic excitations

In Landau's formalism of phase transition theory [59], all phases of matter can be characterized by a local order parameter, and the phase transition between different phases corresponds to a symmetry breaking. The discovery of the integer and fractional quantum Hall effects enabled the realization of a new phase of matter beyond Landau's paradigm. This new phase should be described by topological invariants and was dubbed the topological phase. The topological invariants remain unchanged under the adiabatic deformations of the system as long as the bulk gap is not closed. Topological invariants arise from geometrical phases, as we show next.

3.1 Geometric phases, Zak phase, and Chern number

Whenever a quantum system undergoes a cyclic evolution governed by a slow change of parameters, it acquires a phase factor: the geometric phase. Unlike dynamical phases, geometrical phases are not attributed to forces applied to quantum systems. Instead, they are associated with the connection of space [60]. The Aharonov-Bohm phase [61] is a particular case of the geometric phase. It is the relative phase acquired by two electronic wavepackets encircling a magnetic field confined to a solenoid, so that along their paths, the magnetic field is zero, but the vector potential is nonzero. The Aharonov-Bohm phase affects their interference pattern when closing the loop and is proportional to the enclosed magnetic flux. The Aharonov-Bohm phase is topological. As such, it does not depend on the shape or the geometric properties of the particle path, but only on its topological invariants, provided that the particle is moving in a field-free region.

The general form of the geometric phase, the Berry phase, was introduced by Michael Berry [62], who showed that geometric phases arise purely from the adiabatic evolution of a quantum state $|n(\mathbf{R})\rangle$ upon the modification of the parameters \mathbf{R} . Geometrical phases are related to the topological structure formed by the Hilbert space of a quantum system with a Hamiltonian \mathcal{H} , and the space of its adiabatically varied parameter \mathbf{R} , \mathcal{M} . The evolution of the state $|n(\mathbf{R})\rangle$ is tied to the Berry connection, a generalization of the vector potential $\mathbf{A} = i\langle n(\mathbf{R}) | \nabla_{\mathbf{R}} | n(\mathbf{R}) \rangle$. \mathbf{A} gives rise to the Berry phase, γ , after integration along a closed path. Connections allow the differentiation of wavefunctions over \mathbf{R} by providing a unique way of dragging them from one Hilbert space in $\mathcal{M} \times \mathcal{H}$ to another. This process is known as parallel transport and can occur provided that a smooth path between both spaces is provided.

For electrons in crystalline systems, the Bloch theorem is applicable regardless of the form of the potential. The Bloch theorem states that the energy eigenstates of an underlying periodic Hamiltonian, known as Bloch states, are given by the wavefunction $|\psi_{n,\mathbf{k}}(\mathbf{r})\rangle = e^{i\mathbf{k}\cdot\mathbf{r}}|u_{n,\mathbf{k}}(\mathbf{r})\rangle$. Here, n denotes the energy band index, \mathbf{k} is the electronic quasi-momentum, and $u(\mathbf{r})$ is a periodic function with the periodicity of the underlying Bravais lattice vector \mathbf{R} , $u(\mathbf{r} + \mathbf{R}) = u(\mathbf{r})$. Consider an isolated band. Following $\mathbf{k} = \mathbf{k}(\mathbf{t})$ throughout the Brillouin zone, the Bloch state of an electron traverses a closed path in the reciprocal quasi-momentum parameter space, and a geometric phase emerges. Now, space of parameter \mathbf{R} corresponds to the adiabatic evolution of the original Bloch state $|\psi_{n,\mathbf{k}}(\mathbf{r}, t_0)\rangle$ through a closed loop in the first Brillouin zone [63]. This results in the geometric (Berry) phase, $\gamma = \int \mathbf{A}(\mathbf{k}) \cdot d\mathbf{k}$, where \mathbf{A} is the Berry connection for a given energy band, $-i\langle u_{\mathbf{k}} | \nabla_{\mathbf{k}} | u_{\mathbf{k}} \rangle$, a gauge-dependent parameter that does not correspond to an observable. Nevertheless, the geometric phase is gauge invariant. The Berry curvature, defined as the curl of the Berry connection $\mathbf{\Omega}(\mathbf{k}) = \nabla_{\mathbf{k}} \times \mathbf{A}(\mathbf{k})$, is a measure of the local rotation of the electronic wavepacket as it transverses the Brillouin zone and like the connection is gauge dependent. Through the Stokes theorem, one can write the geometric phase as an integral over the manifold of the Berry curvature. If such a manifold is closed (a torus, for instance), then the result is topological and quantized by a 2π multiple of the topological index named Chern number.

The Berry phase presented above applies to generic systems described by a parameter-dependent Hamiltonian. It is a topological invariant that can be used to judge whether or not the system is a first-order topological insulator.

If the Berry phase is computed over a noncontractible loop of the Brillouin zone torus, it is known as the Zak phase, $\mathcal{Z} = \int_L d\mathbf{k} \cdot \mathbf{A}_n(\mathbf{k}) \pmod{2\pi}$. Zak [64] introduced

the concept of geometric phase for Bloch electrons in one-dimensional periodic lattices in 1989. Zak phase is useful for identifying the topological phase in a one-dimensional system, where the integral interval goes from 0 to $2\pi/a$, with a the lattice constant. When the one-dimensional system has inversion symmetry, the nonzero Zak phase indicates that the system is in a topological phase. Besides the Berry phase, the Chern number is another typical topological invariant adopted to describe first-order topological insulators. It is expressed as the integration of the Berry curvature in the Brillouin zone $C_n = \frac{1}{2\pi} \int_{BZ} d^2k \Omega_{n,z}(\mathbf{k})$ [63].

The Chern number is an integer. For $C_n = 0$, the system is in a trivial phase. If the system supports a nontrivial topological phase, then C_n is a nonzero integer [65]. For time-reversal symmetric systems, the Chern number is always zero. To characterize first-order topological insulator in time-reversal symmetric systems, the Z_2 invariant is often used [63]. For higher-order topological insulators, typical topological invariants include bulk polarization and \mathbb{Z}_N Berry phases.

The definition of the Chern number presented above is used in fermionic systems. Because magnons are bosons, the expression of C_n must be adapted for magnetic systems. The Hamiltonian of bosons can be expressed in terms of a para-unitary matrix T_k instead of a unitary matrix. A projection operator in the vector space can be defined in terms of T_k , $Q_n = T_k \Gamma_n \sigma_3 T_k^\dagger \sigma_3$, where Γ_n is a diagonal matrix taking +1 for the n th diagonal component and zero otherwise, and a diagonal matrix σ_3 takes +1 in the particle space while -1 in the hole space. Then, the Berry curvature for bosons is given by $B_n = i\epsilon_{\mu\nu} \text{Tr}(Q_n (\partial_{k_\mu} Q_n) (\partial_{k_\nu} Q_n))$, and the Chern number in magnonic systems becomes $C_n = \frac{1}{2\pi} \int_{BZ} d^2k B_n(\mathbf{k})$ [63, 66].

3.2 Spin waves and magnons

The most peculiar characteristic of topological phases, like the topological insulator [67], is that they can support chiral edge (in 2D lattices) and surface (in 3D lattices) states, which are absent in conventional insulators. The topological edge and surface states are wave modes that are confined at the boundary/surface of the system and generally have a specific chirality. These properties are topologically protected, enabling them to be immune to moderate disorders and defects.

The bulk-boundary correspondence dictates that the bulk property of topological insulators [68] determines the character of edge or surface modes. A first-order m -dimensional topological insulator has $(m - 1)$ -dimensional topological edge/surface states. Higher-order m -dimensional k -order topological insulators, on the other hand, allow for $(m - k)$ -dimensional topological boundary modes ($2 \leq k \leq n$), such as corner states and hinge states.

In the magnetic arena, there has been intensive research on topological magnetic states in the last decade. Magnons, quanta of spin waves, propagate localized spin fluctuations in solids. Research associated to the detection, and manipulation of magnons is often called magnonics. Topologically protected unidirectional surface spin waves present advantages over their trivial counterparts because they are robust against internal and external perturbations. The study of the topological properties of magnons began in 2010 with the experimental observation of the magnon Hall effect (MHE) in the insulating pyrochlore ferromagnet $\text{Lu}_2\text{V}_2\text{O}_7$ [69]. The topological transport of charge-neutral excitations in insulators produces a thermal analog of the topological Hall effect by electrons. A model of uncompensated net magnon edge currents was proposed to explain the emerging MHE where the longitudinal

temperature gradient induces a transverse thermal current, [66, 70]. The concept of a topological magnon insulator was then adopted to describe a large class of systems that support chiral magnon current circulating around their boundaries. The MHE was subsequently observed in garnet magnets (for example, yttrium iron garnets), kagome and pyrochlore magnets, and frustrated pyrochlore quantum magnets [66, 68].

Topological bands also arise in gapless magnetic systems. This is the case of topological semimetals, which include the Dirac and Weyl semimetals. Weyl semimetals are characterized by degenerate points, denoted Weyl points resulting from the linear crossing of two bands. Weyl points come in pairs with opposite chiralities. The band inversion happens between two paired Weyl nodes, leading to the generation of topologically protected Fermi-arc-like surface states. Weyl semimetals can also support higher-order topological edge states [71] and emerge when at least one of the two, that is, time reversal symmetry or inversion symmetry, is broken. On the other hand, Dirac points with fourfold degenerate band touchings characterize Dirac semimetals. Dirac semimetals respect both the time-reversal and inversion symmetries, while their stability requires additional crystalline symmetries, for example, the rotational.

Besides the topological magnon insulator and the Dirac and Weyl semimetals, there are other sources of topological effect in magnon systems, as is the case of the topological magnon polarons resulting from the spin-lattice coupling and the topological phases of magnon Bogoliubov-de Gennes systems [68]. Nowadays, the search for topological magnons in antiferromagnet and ferrimagnet is an active field of research [72].

The experimental and theoretical advances on topological magnons over the past decade have focused on finding the topological magnon insulator state in lattices with triangular motifs such as the kagome (or pyrochlore) and honeycomb lattices. Simultaneously, different microscopic mechanisms leading to nontrivial topology have been identified for such lattices. Examples include the Dzyaloshinskii-Moriya (DM) interaction due to the inversion symmetry broken, the magnetic dipolar interaction, the pseudodipolar exchange interaction, and internal fields from intrinsic magnetic textures.

3.3 Topological transport with magnons: The Magnon Hall effect

The MHE was observed experimentally in the insulating pyrochlore ferromagnet $\text{Lu}_2\text{V}_2\text{O}_7$ [69]. In the experiment, a transverse heat current was observed when a temperature gradient was applied longitudinally. The experiment showed two peculiarities: i) the thermal Hall conductivity steeply increased and saturated in the low magnetic field regime. This behavior could not be explained by the normal Hall effect where the conductivity is proportional to the magnetic field strength. It resembles instead the anomalous Hall effect [73] due to spontaneous magnetization. ii) the thermal Hall conductivity decreased in the high field regime, which could not be attributed to phonons. It was concluded then that the transverse heat current was due to the MHE. The model included the ferromagnetic exchange interaction and a non-zero DM interaction induced by the spin-orbit coupling, which breaks the inversion symmetry. Subsequently, it was demonstrated that a magnon wave packet subjected to a temperature gradient acquires an anomalous velocity perpendicular to the gradient, which is associated with the magnon edge currents [74].

The thermal Hall conductivity originates from the Berry curvature $\Omega_n(\mathbf{k})$ in momentum space. Therefore, the transverse thermal Hall conductivity κ^{xy} is given by

$$\kappa^{xy} = -\frac{k_B^2 T}{\hbar v} \sum_{n, \mathbf{k}} c_2(\rho_n) \Omega_{n, \mathbf{z}}(\mathbf{k}) \quad (14)$$

where $\rho_n[\varepsilon_n(\mathbf{k})]$ is the Bose distribution function, k_B is the Boltzman constant, v is the volume, and T is the temperature. $c_2(\rho_n) = (1 + \rho) \left(\log \frac{1+\rho}{\rho} \right)^2 - (\log \rho)^2 - 2\text{Li}_2(-\rho)$, with $\text{Li}_2(z)$ the polylogarithm function. In the MHE, when the system is in equilibrium, the edge magnon currents exist due to the confining potential, and they circulate along the boundary. The currents are equal at the two edges of a sample leading to a vanishing thermal current through the magnet. If a temperature gradient is applied, the magnons will flow from the high-temperature region to the low-temperature region, which breaks the balance of the heat current in the two opposite edges, leading to a finite thermal Hall current. Such edge magnon currents are spin waves chiral edge states resulting from the nontrivial topology of magnon bands. The one-way chiral edge transport is topologically immune to defects and disorder. Since the discovery of MHE in pyrochlore ferromagnetic insulators, the same effect has been observed in other magnetic materials. Examples include the frustrated pyrochlore quantum magnet $\text{Tb}_2\text{Ti}_2\text{O}_7$, YIG, and the kagome magnet $\text{Cu}(1, 3 - \text{bcd})$ [63].

Noncollinear magnetic textures like skyrmions can generate a fictitious magnetic field, leading to the magnon thermal Hall effect. This Hall effect is due to the nonzero topological charge of magnetic texture and is called the topological magnon Hall effect (TMHE) [75].

Over the past decade, much effort has been devoted to studying the topological properties of magnons. Magnetic systems and models include a plethora of materials, lattices, interfaces, and heterostructures. Next, we present a few examples of topological magnon insulators and semimetals.

3.4 Magnetic topological matter: Magnon insulators, Dirac, and Weyl semimetals

3.4.1 Topological magnon insulators

Chern number is a property of the bulk of a system and determines the propagation direction and the number of topologically nontrivial edge modes. The sum of Chern numbers up to the n th band $\nu_n = \sum_{j \leq n} C_n$ is the winding number of the edge states in bandgap n . ν_n corresponds to the number of topologically nontrivial edge states in the n th bandgap and $\text{sign}(\nu_n)$ determines their propagation direction.

The topological magnon insulator was predicted in the ferromagnetic kagome lattice with Heisenberg and Dzyaloshinskii-Moriya (DM) interactions [68]. The ferromagnetic kagome lattice allows four topologically different phases by tuning the parameters J_{NN}/J_N and D/J_N , where, in this context, J_N and J_{NN} denote nearest-neighbor and next-nearest-neighbor (NNN) symmetric exchange interaction and D is the nearest-neighbor antisymmetric interaction. In this case, the nontrivial topology of magnon in the kagome lattice is brought about by the strong spin-orbit coupling.

The magnetic dipolar interaction can also endow spin wave volume modes with nonzero Chern number. Recently, it has been found theoretically that dipolar zig-zag chains, one-dimensional lattices of point dipoles with a two-point basis, are building blocks of two-dimensional lattices with topological magnon bands. The system is a two-band model where the only coupling between spins is due to the long-ranged

dipolar interactions. Dipolar coupling brings about topology by locking the spin's degrees of freedom to the lattice, giving rise to an effective spin-orbit interaction. Stripes, built from a finite number of chains, host topological chiral edge states. Tuning the vertical distance between zig-zag chains allows for the direct control of magnonic frequencies, the velocity of the chiral edge modes, and the transverse thermal conductivity. The Chern numbers of the two spin-wave volume bands are exchanged for a critical vertical distance due to two band touchings, which yield two monopoles and endow the Berry phase with a divergence. This triggers the exchange of the Chern numbers between the bands. Such topological phase transition causes the change of sign of both the Hall conductivity and the sense of motion of edge states in the system [76].

In artificial systems, the observation of topological insulator states has been recently realized in a tunable superconducting qubit chain, which exhibits both nontrivial topological invariants and topological edge states [63].

3.4.2 Dirac and Weyl semimetals

The Heisenberg magnet on the honeycomb lattice exhibits Dirac points. The Hamiltonian can be expressed as $H = - \sum J_{ij} \mathbf{S}^A \cdot \mathbf{S}^B$, where the summation runs over nearest neighbors, \mathbf{S}^A and \mathbf{S}^B are the spins for two different sublattices A and B , and J_{ij} is the exchange constant. Around the band degeneracy points K and K' , the dispersion relation is linear. If the system is an antiferromagnetic honeycomb lattice, the energy dispersion degenerates only at the Γ point, but the dispersion remains linear. For magnons, the time-reversal operator can be defined for the Bogoliubov Hamiltonian. As long as the pseudo spin time-reversal symmetry is preserved, the Dirac points exist in the Brillouin zone, but if the time-reversal symmetry is broken, a gap will open at the Dirac points, leading to a topological insulator. If the Hamiltonian only contains NN exchange interaction, the magnon band structure is gapless for the honeycomb ferromagnets, and if a NNN exchange interaction is considered, it shifts the positions of the Dirac points. However, if a DM interaction is introduced, the pseudo spin time-reversal symmetry of the system is broken, and a gap opens at the Dirac points. Chromium trihalides CrX_3 ($X = \text{F, Cl, Br, and I}$) are a practical example of ferromagnets consisting of van der Waals-bonded stacks of honeycomb layers, which display two spin-wave modes with energy dispersion similar to that for the electrons in graphene. The gap at the Dirac points in CrI_3 was observed experimentally [68] by using inelastic neutron scattering. The observation of a sizeable spin-wave gap indicates that the spin-orbit coupling plays a vital role in the physics of topological spin excitations not only in bulk honeycomb ferromagnet CrI_3 but also in its monolayers. By using inelastic neutron scattering method, Dirac magnons have also been observed in the three-dimensional quantum XY magnet CoTiO_3 . A gap arises from the bond-anisotropic exchange coupling, due to quantum order by disorder, which pins the order parameter to the crystal axes [77].

The magnon bands in a magnonic Weyl semimetal are crossed in pairs at the points dubbed Weyl nodes. The Weyl nodes are monopoles of Berry curvature and are characterized by the integer topological charge, or chirality [68]. Because the net topological charges in the entire Brillouin zone must be zero, the Weyl nodes appear in pairs with opposite topological charges of ± 1 . In Weyl semimetals, the topologically protected chiral surface states between each pair of Weyl nodes exist on the system surfaces, and the equal energy contour of these surface states forms arcs, with the arc

number equaling the number of paired Weyl nodes. The vital hallmark of magnonic Weyl semimetal is the magnon arcs on system surfaces. The pyrochlore ($\text{Lu}_2\text{V}_2\text{O}_7$) ferromagnet with DM interaction is an intrinsic magnonic Weyl semimetal with the pair of Weyl nodes having opposite topological charges. In the model of such a system, the effective spin Hamiltonian includes NN exchange interaction, NN DM interaction, and Zeeman interaction. The magnonic chiral anomaly could be realized by applying inhomogeneous electric and magnetic fields perpendicular to each other: when a magnon moves its magnetic moment can interact with the electric field and it acquires an AC phase, while the magnetic field is used to drive the magnon flow. The field drives magnons to move from one Weyl node to the other through the zeroth magnonic Landau level and results in the imbalance of chirality, which is the signature of magnonic chiral anomaly [78, 79].

4. Conclusion

Topology is a wide and deep-rooted concept in physics. It manifests as defects in the order parameter, topological order in the ground state, or in the spectrum of excitation of ordered phases. Tools from topology and homotopy theory have proven indispensable for the prediction of a variety of new phenomena. In this chapter, we have focused on the topological properties in ordered media such as magnetic textures and low energy spin fluctuations.

We have introduced basic elements of homotopy theory to classify self-localized magnetic textures in two and three dimensions. As examples, we considered magnetic domain walls and skyrmions in two dimensions.

The complex structures of domain walls can be understood in terms of elementary topological defects, with winding numbers that determine the switching processes in nanomagnets. Skyrmions emerge as a single entity but also in ordered arrays. The dynamics of skyrmions, as a response to external forces from electric currents, thermal gradients, and magnetic fields, realizes deflected trajectories in analogy to the Magnus effect. The dynamics of skyrmions is determined by the Landau-Lifschitz-Gilbert equation, which, through reduction of degrees of freedoms, is simplified to Thiele's equation. This framework embodies a rigid single-skyrmion dynamics, which can be generalized in multiple ways, in particular to include size deformations leading to inertial motions.

Spin waves, propagating excitations in magnetic materials, manifest too topological characteristics. This time the topology is rooted in the band spectrum of the system's linear excitations rather than in the order parameter. The Chern number is one of the topological invariants that help out to characterize the topological character of the bands. Topological spin wave bands are usually accompanied with exotic phenomena such as the emergence of chiral edge states and the magnon Hall effect. Research on topological band theory constitutes the central tool for studying either bosonic or fermionic topological phases of matter.

Acknowledgements

P.M. thanks support from Fondecyt under Grant No. 1210083. R.T. would like to thank Sebastian Diaz for the preparation of images and fruitful discussions.

Conflict of interest

The authors declare no conflict of interest.


Author details

Paula Mellado^{*†} and Roberto E. Troncoso[†]
Universidad Adolfo Ibáñez, Santiago, Chile

*Address all correspondence to: pmellado93@gmail.com

† These authors contributed equally.

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Moessner R, Moore JE. Topological Phases of Matter. Cambridge, United Kingdom: Cambridge University Press; 2021
- [2] Dzyaloshinsky I. A thermodynamic theory of “weak” ferromagnetism of antiferromagnetics. *Journal of Physics and Chemistry of Solids*. 1958;**4**(4): 241-255
- [3] Moriya T. New mechanism of anisotropic superexchange interaction. *Physical Review Letters*. 1960;**4**(5):228
- [4] Debell K, Macisaac A, Whitehead J. Dipolar effects in magnetic thin films and quasi-two-dimensional systems. *Reviews of Modern Physics*. 2000;**72**(1):225
- [5] Mermin ND. The topological theory of defects in ordered media. *Reviews of Modern Physics*. 1979;**51**:591-648
- [6] St H. Homotopy Theory. Cambridge, Massachusetts, United States: Academic Press; 1959
- [7] Thiaville A, Miltat J. Topology and magnetic domain walls. In: *Topology in Magnetism*. Midtown Manhattan, New York City: Springer; 2018. pp. 41-73
- [8] Nakahara M. *Geometry, Topology and Physics*. Boca Raton, Florida: CRC Press; 2018
- [9] Skyrme THR. A unified field theory of mesons and baryons. *Nuclear Physics*. 1962;**31**:556-569
- [10] Nagaosa N, Tokura Y. Topological properties and dynamics of magnetic skyrmions. *Nature Nanotechnology*. 2013;**8**:899-911
- [11] Zarzuela R, Ochoa H, Tserkovnyak Y. Hydrodynamics of three-dimensional skyrmions in frustrated magnets. *Physical Review B*. 2019 Aug;**100**:054426
- [12] Göbel B, Mertig I, Tretiakov OA. Beyond skyrmions: Review and perspectives of alternative magnetic quasiparticles. *Physics Reports*. 2021; **895**:1-28
- [13] Wang XS, Qaiumzadeh A, Brataas A. Current-driven dynamics of magnetic Hopfions. *Physical Review Letters*. 2019; **123**:147203
- [14] Kosterlitz JM. Kosterlitz–Thouless physics: A review of key issues. *Reports on Progress in Physics*. 2016;**79**(2): 026001
- [15] Beg M, Pepper RA, Cortés-Ortuño D, Atie B, Bisotti MA, Downing G, et al. Stable and manipulable Bloch point. *Scientific Reports*. 2019;**9**(1):1-8
- [16] Mellado P, Petrova O, Shen Y, Tchernyshyov O. Dynamics of magnetic charges in artificial spin ice. *Physical Review Letters*. 2010;**105**(18):187206
- [17] Cervera CJG. *Magnetic Domains and Magnetic Domain Walls*. New York, NY, USA: New York University; 1999
- [18] Kläui M. Head-to-head domain walls in magnetic nanostructures. *Journal of Physics. Condensed Matter*. 2008; **20**(31):313001
- [19] LaBonte A. Two-dimensional Bloch-type domain walls in ferromagnetic films. *Journal of Applied Physics*. 1969; **40**(6):2450-2458
- [20] Tchernyshyov O, Chern GW. Fractional vortices and composite domain walls in flat nanomagnets.

- Physical Review Letters. 2005;**95**(19): 197204
- [21] Chern GW, Youk H, Tchernyshyov O. Topological defects in flat nanomagnets: The magnetostatic limit. *Journal of Applied Physics*. 2006; **99**(8):08Q505
- [22] Youk H, Chern GW, Merit K, Oppenheimer B, Tchernyshyov O. Composite domain walls in flat nanomagnets: The magnetostatic limit. *Journal of Applied Physics*. 2006;**99**(8): 08B101
- [23] Backes D, Schieback C, Kläui M, Junginger F, Ehrke H, Nielaba P, et al. Transverse domain walls in nanoconstrictions. *Applied Physics Letters*. 2007;**91**(11):112502
- [24] Tretiakov O, Clarke D, Chern GW, Bazaliy YB, Tchernyshyov O. Dynamics of domain walls in magnetic nanostrips. *Physical Review Letters*. 2008;**100**(12): 127204
- [25] Mougin A, Cormier M, Adam J, Metaxas P, Ferré J. Domain wall mobility, stability and Walker breakdown in magnetic nanowires. *EPL (Europhysics Letters)*. 2007;**78**(5): 57007
- [26] Ghosh A, Huang KS, Tchernyshyov O. Annihilation of domain walls in a ferromagnetic wire. *Physical Review B*. 2017;**95**(18):180408
- [27] Zhu F, Chern G, Tchernyshyov O, Zhu X, Zhu J, Chien C. Magnetic bistability and controllable reversal of asymmetric ferromagnetic nanorings. *Physical Review Letters*. 2006;**96**(2): 027205
- [28] Bogdanov AN, Rößler UK, Shestakov AA. Skyrmions in nematic liquid crystals. *Physical Review E*. 2003; **67**:016602
- [29] Fert A, Cros V, Sampaio J. Skyrmions on the track. *Nature Nanotechnology*. 2013;**8**(3):152-156
- [30] Jonietz F, Mühlbauer S, Pfleiderer C, Neubauer A, Münzer W, Bauer A, et al. Spin transfer torques in MnSi at ultralow current densities. *Science*. 2010; **330**(6011):1648-1651
- [31] Mühlbauer S, Binz B, Jonietz F, Pfleiderer C, Rosch A, Neubauer A, et al. Skyrmion lattice in a chiral magnet. *Science*. 2009;**323**(5916):915-919
- [32] Neubauer A, Pfleiderer C, Binz B, Rosch A, Ritz R, Niklowitz PG, et al. Topological hall effect in the a phase of MnSi. *Physical Review Letters* 2009 102: 186602
- [33] Münzer W, Neubauer A, Adams T, Mühlbauer S, Franz C, Jonietz F, et al. Skyrmion lattice in the doped semiconductor $\text{Fe}_{1-x}\text{Co}_x\text{Si}$. *Physical Review B*. 2010;**81**:041203
- [34] Milde P, Köhler D, Seidel J, Eng LM, Bauer A, Chacon A, et al. Unwinding of a Skyrmion lattice by magnetic monopoles. *Science*. 2013;**340**(6136):1076-1080
- [35] Yu XZ, Onose Y, Kanazawa N, Park JH, Han JH, Matsui Y, et al. Real-space observation of a two-dimensional skyrmion crystal. *Nature*. 2010; **465**(7300):901-904
- [36] Shibata K, Yu XZ, Hara T, Morikawa D, Kanazawa N, Kimoto K, et al. Towards control of the size and helicity of skyrmions in helimagnetic alloys by spin-orbit coupling. *Nature Nanotechnology*. 2013;**8**(10):723-728
- [37] Yu XZ, Kanazawa N, Onose Y, Kimoto K, Zhang WZ, Ishiwata S, et al.

Near room-temperature formation of a skyrmion crystal in thin-films of the helimagnet FeGe. *Nature Materials*. 2010;**10**(2):106-109

[38] Nagao M, So YG, Yoshida H, Isobe M, Hara T, Ishizuka K, et al. Direct observation and dynamics of spontaneous skyrmion-like magnetic domains in a ferromagnet. *Nature Nanotechnology*. 2013;**8**(5): 325-328

[39] Seki S, Yu XZ, Ishiwata S, Tokura Y. Observation of Skyrmions in a Multiferroic material. *Science*. 2012; **336**(6078):198-201

[40] Heinze S, von Bergmann K, Menzel M, Brede J, Kubetzka A, Wiesendanger R, et al. Spontaneous atomic-scale magnetic skyrmion lattice in two dimensions. *Nature Physics*. 2011; **7**(9):713-718

[41] Heinrich B. Skyrmion birth at the notch. *Nature Nanotechnology*. 2021; **16**(10):1051-1051

[42] Berger L. Motion of a magnetic domain wall traversed by fast-rising current pulses. *Journal of Applied Physics*. 1992;**71**(6): 2721-2726

[43] Mochizuki M, Yu XZ, Seki S, Kanazawa N, Koshibae W, Zang J, et al. Thermally driven ratchet motion of a skyrmion microcrystal and topological magnon Hall effect. *Nature Materials*. 2014;**13**(3):241-246

[44] Troncso RE, Núñez AS. Thermally assisted current-driven skyrmion motion. *Physical Review B*. 2014;**89**: 224403

[45] Lin SZ, Batista CD, Reichhardt C, Saxena A. ac current generation in chiral magnetic insulators and Skyrmion

motion induced by the spin Seebeck effect. *Physical Review Letters*. 2014; **112**:187203

[46] Thiele AA. Steady-state motion of magnetic domains. *Physical Review Letters*. 1973;**30**:230-233

[47] Brown WF. Thermal fluctuations of a single-domain particle. *Physics Review*. 1963;**130**:1677-1686

[48] Kubo R, Hashitsume N. Brownian motion of spins. *Progress of Theoretical Physics Supplement*. 1970;**06**(46): 210-220

[49] Duine RA, Núñez AS, MacDonald AH. Thermally assisted current-driven Domain-Wall motion. *Physical Review Letters*. 2007; **98**:056605

[50] Slonczewski JC. Current-driven excitation of magnetic multilayers. *Journal of Magnetism and Magnetic Materials*. 1996;**159**(1):L1-L7

[51] Tatara G, Kohno H. Theory of current-driven Domain Wall motion: Spin transfer versus momentum transfer. *Physical Review Letters*. 2004; **92**:086601

[52] Xiao J, Zangwill A, Stiles MD. Spin-transfer torque for continuously variable magnetization. *Physical Review B*. 2006; **73**:054428

[53] Zhang S, Li Z. Roles of nonequilibrium conduction electrons on the magnetization dynamics of Ferromagnets. *Physical Review Letters*. 2004;**93**:127204

[54] Yamanouchi M, Chiba D, Matsukura F, Dietl T, Ohno H. Velocity of domain-wall motion induced by electrical current in the ferromagnetic

semiconductor (Ga,Mn)As. *Physical Review Letters*. 2006;**96**:096601

[55] Yamaguchi A, Ono T, Nasu S, Miyake K, Mibu K, Shinjo T. Real-space observation of current-driven domain wall motion in submicron magnetic wires. *Physical Review Letters*. 2004;**92**:077205

[56] Landeros P, Nunez AS. Domain wall motion on magnetic nanotubes. *Journal of Applied Physics*. 2010;**108**(3):033917

[57] Sutcliffe P. Hopfions in chiral magnets. *Journal of Physics A: Mathematical and Theoretical*. 2018; **51**(37):375401

[58] Raftrey D, Fischer P. Field-driven dynamics of magnetic Hopfions. *Physical Review Letters*. 2021;**127**:257201

[59] Hoffmann KH, Tang Q. Ginzburg-Landau Phase Transition Theory and Superconductivity. Birkhäuser; 2012

[60] Cohen E, Larocque H, Bouchard F, Nejadstatti F, Gefen Y, Karimi E. Geometric phase from Aharonov–Bohm to Pancharatnam–Berry and beyond. *Nature Reviews Physics*. 2019;**1**(7): 437-449

[61] Peshkin M. The Aharonov-Bohm effect part one: Theory. The Aharonov-Bohm Effect. 1989;**1989**:1-34

[62] Berry MV. Quantal phase factors accompanying adiabatic changes. *Proceedings of the Royal Society of London A Mathematical and Physical Sciences*. 1802;**1984**(392):45-57

[63] Wang X, Wang X. Topology in magnetism. Chirality, Magnetism and Magnetoelectricity. 2021;**2021**: 357-403

[64] Zak J. Berry's phase for energy bands in solids. *Physical Review Letters*. 1989; **62**(23):2747

[65] Prodan E. Robustness of the spin-Chern number. *Physical Review B*. 2009; **80**(12):125327

[66] McClarty PA. Topological magnons: A review. *Annual Review of Condensed Matter Physics*. 2022;**13**:171-190

[67] Shen SQ. Topological Insulators. Midtown Manhattan, New York City: Springer; 2012

[68] Li ZX, Cao Y, Yan P. Topological insulators and semimetals in classical magnetic systems. *Physics Reports*. 2021; **915**:1-64

[69] Onose Y, Ideue T, Katsura H, Shiomi Y, Nagaosa N, Tokura Y. Observation of the magnon Hall effect. *Science*. 2010;**329**(5989):297-299

[70] Taguchi Y, Oohara Y, Yoshizawa H, Nagaosa N, Tokura Y. Spin chirality, Berry phase, and anomalous Hall effect in a frustrated ferromagnet. *Science*. 2001;**291**(5513):2573-2576

[71] Wang S, Lin BC, Wang AQ, Yu DP, Liao ZM. Quantum transport in Dirac and Weyl semimetals: A review. *Advances in Physics: X*. 2017;**2**(3): 518-544

[72] Bao S, Wang J, Wang W, Cai Z, Li S, Ma Z, et al. Discovery of coexisting Dirac and triply degenerate magnons in a three-dimensional antiferromagnet. *Nature Communications*. 2018;**9**(1):1-7

[73] Nagaosa N, Sinova J, Onoda S, MacDonald AH, Ong NP. Anomalous hall effect. *Reviews of Modern Physics*. 2010;**82**(2):1539

[74] Mook A, Henk J, Mertig I. Magnon Hall effect and topology in kagome lattices: A theoretical investigation. *Physical Review B*. 2014; **89**(13):134409

[75] van Hoogdalem KA, Tserkovnyak Y, Loss D. Magnetic texture-induced thermal Hall effects. *Physical Review B*. 2013;**87**(2):024402

[76] Mellado P. Intrinsic topological magnons in arrays of magnetic dipoles. *Scientific Reports*. 2022;**12**(1):1-15

[77] Elliot M, McClarty PA, Prabhakaran D, Johnson R, Walker H, Manuel P, et al. Order-by-disorder from bond-dependent exchange and intensity signature of nodal quasiparticles in a honeycomb cobaltate. *Nature Communications*. 2021;**12**(1):1-7

[78] Su Y, Wang X, Wang X. Magnonic Weyl semimetal and chiral anomaly in pyrochlore ferromagnets. *Physical Review B*. 2017;**95**(22):224403

[79] Su Y, Wang X. Chiral anomaly of Weyl magnons in stacked honeycomb ferromagnets. *Physical Review B*. 2017;**96**(10):104437

Mechanical System Kinematics and Dynamics on Differentiable Manifolds

Edward J. Haug and Petar Pavesic

Abstract

Topological and vector space attributes of Euclidean space are employed to create a differentiable manifold structure for holonomic mechanical system kinematics and dynamics. A local kinematic parameterization is presented that establishes the regular configuration space as a differentiable manifold. Topological properties of Euclidean space show that this manifold is naturally partitioned into disjoint, maximal, path connected, singularity free domains of kinematic and dynamic functionality. Using the manifold parameterization, d'Alembert variational equations of mechanical system dynamics yield well-posed ordinary differential equations of motion on these domains, without introducing Lagrange multipliers. Solutions of the differential equations satisfy kinematic configuration, velocity, and acceleration constraints and the variational equations of dynamics. Two examples, one planar and one spatial, are treated using the formulation presented. Solutions obtained are shown to satisfy all three forms of kinematic constraint to within specified error tolerances, using fourth order Runge-Kutta numerical integration methods.

Keywords: differentiable manifold, regular configuration space, mechanical system dynamics, holonomic dynamic system, mechanical system kinematics

1. Introduction

The objective of establishing a rigorous ordinary differential equation (ODE) formulation for mechanical system dynamics of holonomic systems, without Lagrange multipliers and the associated differential-algebraic equations (DAE) of motion, has been intensively pursued during the past two decades. The motivation for this objective is to avoid the complexity of index 3 DAE of mechanical system dynamics [1]. This paper provides a foundation for a Lagrange multiplier-free ODE formulation of mechanical system dynamics on differentiable manifolds, including existence, uniqueness, and numerical solution on maximal, path connected, domains of singularity free kinematic and dynamic functionality. A benefit of this ODE formulation is rigorous implementation of explicit and implicit ODE numerical integration methods with effective error control for dynamic simulation of a broad spectrum of mechanical systems.

To meet the foregoing objective, a kinematics formulation in Euclidean space is presented in Section 2. Topological and vector space properties of Euclidean space that are needed to establish a differentiable manifold structure for mechanical system kinematics and dynamics are summarized. A triple slider-crank mechanism is used to illustrate topological characteristics of its regular configuration space. A local parameterization of mechanical system kinematics in terms of independent generalized coordinates is presented in Section 3, using vector space properties of Euclidean space and multivariable calculus. It is shown that parameterizations on a system of neighborhoods in the regular configuration space define that space as a differentiable manifold. Practical methods for computation of vector and matrix functions employed in the parameterization are presented. Local properties of the regular configuration space are extended in Section 4 to global form on maximal, path connected, singularity free, disjoint components of the regular configuration manifold. This provides a foundation for reliable simulation and control of mechanical systems on maximal domains of functionality. It is further shown that configurations in disjoint components cannot be connected by continuous singularity free paths in configuration space. Time dependence of the configuration manifold parameterization is defined in Section 5 and used in Section 6 to reduce the d'Alembert variational formulation of mechanical system dynamics to manifold-based ODE of motion, without use of Lagrange multipliers. It is shown that the second order ODE initial-value problem obtained has a unique solution that depends continuously on problem data; i.e., it is well posed on maximal domains of functionality.

The ODE formulation of a five degree of freedom spatial spinning top is presented in Section 7.1, using properties of Euler parameters defined in the Appendix. It is shown that accurate solutions of the ODE are obtained using a fourth order Runge-Kutta integration method and that configuration, velocity, and acceleration constraints are satisfied to near computer precision. The triple slider-crank introduced in Section 2.3 is treated in Section 7.2, where it is shown that dynamic response of the system predicted by the ODE of motion remains in singularity free constraint manifold components. Finally, conclusions are presented in Section 8.

2. Euclidean topological vector spaces and differentiable manifolds

Mechanical systems considered are comprised of rigid bodies whose configurations are defined by ngc generalized coordinates $\mathbf{q} \in \mathbb{R}^{ngc}$ that are subject to nhc holonomic constraints (where $nhc < ngc$),

$$\Phi(\mathbf{q}) = \mathbf{0} \quad (1)$$

It is assumed that the vector function $\Phi(\mathbf{q}) = [\Phi_1(\mathbf{q}), \dots, \Phi_{nhc}(\mathbf{q})]^T$ has as many continuous derivatives as needed in the analysis that follows. While the development herein is carried out with constraints that do not depend explicitly on time (scleronic constraints), the formulation can be extended to time dependent (rheonomic) constraints, with additional terms appearing due to derivatives with respect to time.

2.1 Euclidean topological space

Two related structures are considered on the set

$$\mathbb{R}^n = \left\{ \mathbf{q} = [q_1, \dots, q_n]^T, q \in \mathbb{R} \right\}$$

in which kinematics and dynamics of mechanical systems take place. One is the structure of a vector space where addition of elements and multiplication by scalars are defined component-wise, i.e., for $\mathbf{q}, \mathbf{q}' \in \mathbb{R}^n$ and $\alpha \in \mathbb{R}$, define

$$\begin{aligned} \mathbf{q} + \mathbf{q}' &= [q_1 + q'_1, \dots, q_n + q'_n]^T \\ \alpha \mathbf{q} &= [\alpha q_1, \dots, \alpha q_n]^T \end{aligned}$$

As a vector space, \mathbb{R}^n is n -dimensional in the sense that given a linearly independent collection of elements $\mathbf{q}^1, \dots, \mathbf{q}^n \in \mathbb{R}^n$, every element $\mathbf{q} \in \mathbb{R}^n$ can be written uniquely as a linear combination of the form $\mathbf{q} = \alpha_1 \mathbf{q}^1 + \dots + \alpha_n \mathbf{q}^n$. There is also a standard *scalar product* on \mathbb{R}^n given as

$$\langle \mathbf{q}, \mathbf{q}' \rangle := \mathbf{q}^T \mathbf{q}' = \sum_{i=1}^n q_i q'_i.$$

As usual, from a scalar product one can define a *norm* $\|\mathbf{q}\| := \sqrt{\langle \mathbf{q}, \mathbf{q} \rangle}$ of a vector \mathbf{q} and the *Euclidean distance* $d(\mathbf{q}, \mathbf{q}') := \|\mathbf{q} - \mathbf{q}'\|$ between vectors \mathbf{q} and \mathbf{q}' .

Although the distance function d is easy to compute and it makes \mathbb{R}^n into a metric space, it is not very useful for the study of subspaces of \mathbb{R}^n . Consider for example two points on a spiral line in a plane; the straight-line Euclidean distance between them completely ignores the geometry of the spiral and is very different from the natural (or intrinsic) distance between the points, which is measured along the spiral. The intrinsic distance is, on the other hand, usually hard to compute as it involves finding geodesic lines between the points. Fortunately, there is a way to avoid these problems by considering the underlying topology on \mathbb{R}^n and its subspaces.

Recall that one defines a *topology* on a set X by declaring some subsets of X to be *open*. This is done in such a way as to satisfy two simple requirements:

1. an arbitrary union of open sets must be open, and
2. a finite intersection of open sets must be open.

A *topological space* is thus a set X with a collection of open sets satisfying conditions (1) and (2). A subset A of a topological space naturally inherits the topology by taking as open subsets in A the intersection of A with open subsets of X . We will say that A with the inherited topology is a *topological subspace* of X .

It is easy to describe continuous maps between two topological spaces X and Y : a function $f : X \rightarrow Y$ is *continuous* if for every open set $U \subseteq Y$ the set $f^{-1}(U) = \{x \in X; f(x) \in U\}$ is an open subset of X . Moreover, a bijective map $f : X \rightarrow Y$ is called a *homeomorphism* if both f and its inverse f^{-1} are continuous. Thus, a homeomorphism between topological spaces X and Y determines a bijection between open sets in X and open sets in Y .

The topologies of relevance for the purposes of mechanical system kinematics are those that are defined by taking as open sets arbitrary unions of open balls with respect to some distance function d . The crucial point here is that different distance functions often generate the same topology. Indeed, in most cases the Euclidean

distance and the intrinsic distance on a smooth submanifold $M \subset \mathbb{R}^n$ generate the same topology on M .

There are two topological properties that are important for kinematics and dynamics of mechanical systems:

1. A topological space is *compact* if every collection of open sets whose union contains the space can be reduced to a finite sub-collection whose union also contains the entire space. This abstract condition becomes much simpler in Euclidean topological spaces, because a subspace of \mathbb{R}^n is compact if it has a finite diameter and if it is closed; i.e., its complement is open in \mathbb{R}^n .
2. A topological space X is *path-connected* if for any two points $x, x' \in X$ there is a continuous path $\omega : [0, 1] \rightarrow X$ starting at $\omega(0) = x$ and ending at $\omega(1) = x'$. In general, one can define a *path-component* of a space X as a maximal path-connected subspace of X . Every topological space can be decomposed in a unique way as a disjoint union of its path-components.

2.2 Differentiable manifolds

Differentiable manifolds are topological spaces in which one can do differential calculus in a similar manner as in Euclidean spaces. This is achieved by covering a topological space M by open subsets homeomorphic to \mathbb{R}^n called ‘charts’, in such a way that whenever two charts overlap there is a transition map between them that is sufficiently differentiable as a map between Euclidean spaces.

A more precise definition adapted to our needs is as follows (see [2]): Let M be a topological subspace of some Euclidean topological space \mathbb{R}^n . A *chart* on M is an open subset $U \subseteq M$ together with a homeomorphism φ from U to an open set of \mathbb{R}^m . A collection of charts $\{(U_i, \varphi_i) | i = 1, 2, \dots\}$ is a *C^r -differentiable atlas* on M if the charts are compatible in the sense that whenever $U_i \cap U_j \neq \emptyset$, then the *transition map*

$$\varphi_j \circ \varphi_i^{-1} : \varphi_i(U_i \cap U_j) \rightarrow \varphi_j(U_i \cap U_j)$$

is at least r -times continuously differentiable as a map between m -dimensional Euclidean spaces. In order to be able to state and solve second-order ODE, we will require that r is at least 2, but will otherwise simply understand that the transition maps are ‘sufficiently differentiable’. A space M with a differentiable atlas as above is called an *m -dimensional differentiable manifold* (of class C^r). Observe that every open subset of M is also a differentiable manifold (of the same dimension and class).

We will frequently use the following construction of differentiable manifolds, based on the Implicit Function Theorem (see [2]). Let D be an open subset of \mathbb{R}^{m+n} and let $\Phi : D \rightarrow \mathbb{R}^n$ be a differentiable map. A point $\mathbf{q} \in \mathbb{R}^{m+n}$ is *regular* if the Jacobian matrix of Φ at \mathbf{q} has rank equal to n , and is *singular* if the rank of that matrix is less than n . If \mathbf{q} is a singular point of Φ , then $\Phi(\mathbf{q})$ is a *singular value* of Φ , while the remaining values of Φ are *regular values*. Then we have the following basic fact (see [2], Theorem 1.38): If $\mathbf{v} \in \mathbb{R}^n$ is a regular value of Φ , then

$$\Phi^{-1}(\mathbf{v}) = \{\mathbf{q} \in \mathbb{R}^{m+n} | \Phi(\mathbf{q}) = \mathbf{v}\}$$

is an m -dimensional differentiable manifold.

2.3 Mechanical system configuration spaces

The *configuration space* C of a mechanical system is the subset of vectors in \mathbb{R}^{ngc} that satisfy Eq. (1),

$$C = \{\mathbf{q} \in \mathbb{R}^{ngc} | \Phi(\mathbf{q}) = \mathbf{0}\} \quad (2)$$

The configuration space is thus a topological subspace of a Euclidean topological space. Since the constraint equation of Eq. (1) is nonlinear, $\Phi(\mathbf{q}) = \Phi(\mathbf{q}') = \mathbf{0}$ does not imply that $\Phi(\mathbf{q} + \mathbf{q}') = \mathbf{0}$, so C is not in general a vector space. Further, the configuration space may contain configurations at which the $nhc \times ngc$ *constraint Jacobian matrix*,

$$\Phi_{\mathbf{q}}(\mathbf{q}) = \left[\frac{\partial \Phi_i}{\partial q_j}(\mathbf{q}) \right] \quad (3)$$

fails to have full rank. Such configurations are *singular*, from both kinematic and dynamic points of view [3], and must be avoided.

From a kinematics point of view, if $\text{rank}(\Phi_{\mathbf{q}}(\mathbf{q}^0)) < nhc$ at some point $\mathbf{q}^0 \in C$, then in general one cannot find $ngc - nhc$ components of \mathbf{q} that can be used as independent coordinates to represent motion of the system in a neighborhood of \mathbf{q}^0 . In other words, the system does not have $ngc - nhc$ *degrees of freedom* in a neighborhood of \mathbf{q}^0 . From a dynamics point of view, it is not possible to represent dynamics of the system as an ODE in any neighborhood of \mathbf{q}^0 . To avoid these problems, the *regular configuration space* is defined as

$$\tilde{C} = \{\mathbf{q} \in \mathbb{R}^{ngc} | \Phi(\mathbf{q}) = \mathbf{0} \text{ and } \text{rank}(\Phi_{\mathbf{q}}(\mathbf{q})) = nhc\} \subseteq C \quad (4)$$

If the set of singular points of Φ is a closed subset of C , then the regular configuration space is a differentiable manifold of dimension $ngc - nhc$.

2.4 Triple slider-crank mechanism illustration

The planar mechanism shown in **Figure 1** is comprised of three sliders that translate along coordinate axes with generalized coordinates q_1, q_3 and q_4 . A crank (body 2) of unit radius is pivoted in and rotates with generalized coordinate q_2 relative to slider 1. A distance constraint of unit length acts between the point on the x_2' -axis shown on the crank periphery and slider 3. Finally, a distance constraint of length 2 units acts between slider 1 and slider 4. Kinematic constraints for this mechanism are

$$\Phi(\mathbf{q}) = \frac{1}{2} \begin{bmatrix} (q_3 - q_1 - \cos q_2)^2 + \sin^2 q_2 - 1 \\ q_1^2 + q_4^2 - 4 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} (q_3 - q_1)(q_3 - q_1 - 2 \cos q_2) \\ q_1^2 + q_4^2 - 4 \end{bmatrix} = \mathbf{0} \quad (5)$$

and the associated constraint Jacobian is

$$\Phi_{\mathbf{q}}(\mathbf{q}) = \frac{1}{2} \begin{bmatrix} -(q_3 - q_1 - \cos q_2) & (q_3 - q_1) \sin q_2 & (q_3 - q_1 - \cos q_2) & 0 \\ q_1 & 0 & 0 & q_4 \end{bmatrix} \quad (6)$$

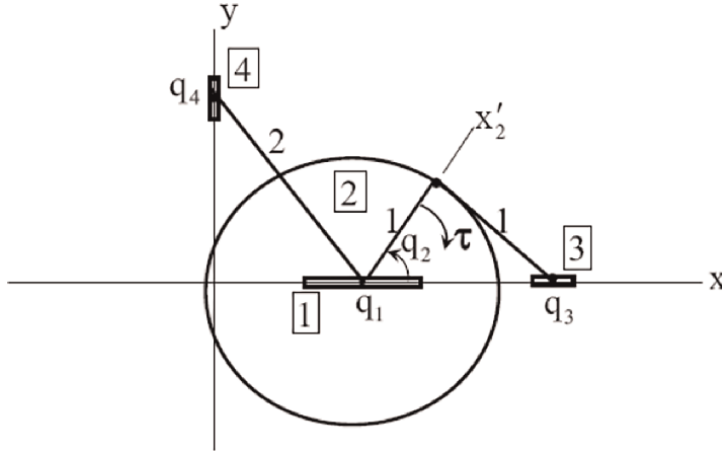


Figure 1.
Triple slider crank mechanism.

Observe that at $q_2 = \pm \frac{\pi}{2}$ we have $q_3 = q_1$ and the entire first row of the matrix is zero, so the Jacobian fails to have full rank. The solution set C of Eq. (5) is the union of the subsets.

$$C_1 = \left\{ \begin{bmatrix} q_1 & q_2 & q_1 + 2 \cos q_2 & \pm \sqrt{4 - q_1^2} \end{bmatrix}^T; -2 \leq q_1 \leq 2, 0 \leq q_2 \leq 2\pi \right\}$$

corresponding to rotations of the crank while sliders 1 and 3 move independently along the axis, and

$$C_2 = \left\{ \begin{bmatrix} q_1 & q_2 & q_1 & \pm \sqrt{4 - q_1^2} \end{bmatrix}^T; -2 \leq q_1 \leq 2, 0 \leq q_2 \leq 2\pi \right\}$$

corresponding to rotations of the crank while the positions of sliders 1 and 3 coincide. Both C_1 and C_2 are path-connected because one can continuously travel between any two positions within each of the sets. However, C_1 and C_2 are not path-components of C , because

$$C_1 \cap C_2 = \left\{ \begin{bmatrix} q_1 & \pm \pi/2 & q_1 & \pm \sqrt{4 - q_1^2} \end{bmatrix}^T; -2 \leq q_1 \leq 2 \right\} \neq \emptyset$$

Note that $C_1 \cap C_2$ is precisely the set of singular points of the map Φ . Thus, in our example the configuration space C is two-dimensional and path-connected, but it contains a one-dimensional subset of singular points where $\text{rank}(\Phi_{\mathbf{q}}(\mathbf{q})) = 1 < n_{hc} = 2$.

By removing singular configurations from C we obtain the regular configuration space \tilde{C} that, as a topological space, splits into four path-components:

$$\begin{aligned} \tilde{C}_1 &= \left\{ \begin{bmatrix} q_1 & q_2 & q_1 & \pm \sqrt{4 - q_1^2} \end{bmatrix}^T; -2 \leq q_1 \leq 2, -\frac{\pi}{2} \leq q_2 \leq \frac{\pi}{2} \right\} \\ \tilde{C}_2 &= \left\{ \begin{bmatrix} q_1 & q_2 & q_1 & \pm \sqrt{4 - q_1^2} \end{bmatrix}^T; -2 \leq q_1 \leq 2, \frac{\pi}{2} \leq q_2 \leq \frac{3\pi}{2} \right\} \\ \tilde{C}_3 &= \left\{ \begin{bmatrix} q_1 & q_2 & q_1 + 2 \cos q_2 & \pm \sqrt{4 - q_1^2} \end{bmatrix}^T; -2 \leq q_1 \leq 2, -\frac{\pi}{2} \leq q_2 \leq \frac{\pi}{2} \right\} \\ \tilde{C}_4 &= \left\{ \begin{bmatrix} q_1 & q_2 & q_1 + 2 \cos q_2 & \pm \sqrt{4 - q_1^2} \end{bmatrix}^T; -2 \leq q_1 \leq 2, \frac{\pi}{2} \leq q_2 \leq \frac{3\pi}{2} \right\} \end{aligned} \quad (7)$$

Within each of these path-components, the Jacobian has maximal rank. Since no rank deficiency of the Jacobian occurs in any of the components of Eq. (7), ODE of motion will be well-posed in each of these path-components (see Section 6). Thus, if the control engineer creates forces and torques that assure the system stays in a path-component, no singularities will be encountered. For example, provided that the torque τ shown in **Figure 1** acting on the crank tends to return the x'_2 -axis fixed in the crank to coincidence with the x -axis and approaches ∞ as q_2 approached $\pm\pi/2$, dynamic simulation that is initiated in any of the path-components of Eq. (7) will remain in that path-component.

3. Kinematic constraint manifolds

Rather than giving an abstract definition of a kinematic constraint manifold, it is instructive to introduce the concept using properties of the nonlinear equations of kinematics.

3.1 Local parameterization of the regular configuration space

By differentiating Eq. (1) with respect to time we obtain

$$\Phi_{\mathbf{q}}(\mathbf{q}^0)\dot{\mathbf{q}},$$

so kinematically admissible velocities $\dot{\mathbf{q}}$ at a point $\mathbf{q}^0 \in \tilde{C}$ are by definition tangent to the regular configuration space at \mathbf{q}^0 .

Let U be the $ngc \times nhc$ matrix $\Phi_{\mathbf{q}}^T(\mathbf{q}^0)$, and let V be a $ngc \times (ngc - nhc)$ matrix whose columns are orthogonal and span the null space of $\Phi_{\mathbf{q}}(\mathbf{q}^0)$ (V may be explicitly computed using singular value decomposition [4]). Since by assumption $\Phi_{\mathbf{q}}(\mathbf{q}^0)$ has maximal rank, the matrix $U^T U$ is positive definite and we have the relations

$$\begin{aligned} U &= \Phi_{\mathbf{q}}^T(\mathbf{q}^0) \\ U^T V &= \mathbf{0} \\ V^T V &= \mathbf{I} \end{aligned} \tag{8}$$

In particular, the columns of U and V span \mathbb{R}^{ngc} .

We may define a parameterization of \tilde{C} in a neighborhood $U^0 \subset \tilde{C}$ of \mathbf{q}^0 as

$$\mathbf{q} = \mathbf{q}^0 + \mathbf{V}\mathbf{v} - \mathbf{U}\mathbf{u} \tag{9}$$

(see Figure 2), which must satisfy

$$\Phi(\mathbf{q}) = \Phi(\mathbf{q}^0 + \mathbf{V}\mathbf{v} - \mathbf{U}\mathbf{u}) = \mathbf{0} \tag{10}$$

We claim that the variables \mathbf{v} and \mathbf{u} comprise a set of *local generalized coordinates* that are in one-to-one correspondence with \mathbf{q} in U^0 . To see this, compute \mathbf{v} and \mathbf{u} using Eqs. (9) and (8) as

$$\begin{aligned}\mathbf{v} &= \mathbf{V}^T(\mathbf{q} - \mathbf{q}^0) \\ \mathbf{u} &= -(\mathbf{U}^T \mathbf{U})^{-1} \mathbf{U}^T(\mathbf{q} - \mathbf{q}^0)\end{aligned}\tag{11}$$

Note that for $\mathbf{q} = \mathbf{q}^0$ we have $\mathbf{v}^0 = \mathbf{0}$ and $\mathbf{u}^0 = \mathbf{0}$.

If we differentiate the function $\Phi(\mathbf{q}^0 + \mathbf{V}\mathbf{v} - \mathbf{U}\mathbf{u})$ with respect to \mathbf{u} we obtain

$$\Phi(\mathbf{q}^0 + \mathbf{V}\mathbf{v} - \mathbf{U}\mathbf{u})_{\mathbf{u}} = -\Phi_{\mathbf{q}}(\mathbf{q})U$$

In particular, $-\Phi_{\mathbf{q}}(\mathbf{q}^0)U = -\mathbf{U}^T U$ is non-singular, because U has maximal rank. This means that \mathbf{u}^0 is a regular point of $\Phi(\mathbf{q}^0 + \mathbf{V}\mathbf{v} - \mathbf{U}\mathbf{u})$, viewed as a function of \mathbf{u} , so we may apply the Implicit Function Theorem to express \mathbf{u} as a function of \mathbf{v} : There exists a neighborhood $V^0 \subseteq \mathbb{R}^{ngc-nhc}$ of $\mathbf{v}^0 = \mathbf{0}$ and a unique (sufficiently) differentiable map $\mathbf{h} : V^0 \rightarrow \mathbb{R}^{nhc}$ such that $\mathbf{h}(\mathbf{v}^0) = \mathbf{u}^0$ and

$$\mathbf{u} = \mathbf{h}(\mathbf{v})$$

is the locally unique solution of Eq. (10), i.e.,

$$\Phi(\mathbf{q}^0 + \mathbf{V}\mathbf{v} - \mathbf{U}\mathbf{h}(\mathbf{v})) = \mathbf{0}\tag{12}$$

for all $\mathbf{v} \in V^0$. In addition, the formula

$$\mathbf{q} = \Psi(\mathbf{v}) := \mathbf{q}^0 + \mathbf{V}\mathbf{v} - \mathbf{U}\mathbf{h}(\mathbf{v})\tag{13}$$

defines a homeomorphism between V^0 and a neighborhood $U^0 := \Psi(V^0) \subset \tilde{C}$ of \mathbf{q}^0 . In other words, (U^0, Ψ^{-1}) is a chart on the manifold \tilde{C} , as described in Section 2.2. By choosing sufficiently many points \mathbf{q}^0 we obtain charts that cover the entire regular configuration space \tilde{C} . These charts are compatible, because the local parametrizations are defined in the neighbourhood of every point by Eq. (13). Note that the matrix V in Eq. (13) is determined by the Jacobian at a given point, up to an orthogonal change of basis, which does not affect differentiability of transition maps. We thus conclude that \tilde{C} admits a differentiable atlas with local parametrizations given as above. If \tilde{C} is compact, then we can find finitely many charts that cover \tilde{C} and thus work with a finite atlas.

Explicit construction of the parametrization $\mathbf{q} = \Psi(\mathbf{v})$ of the regular configuration space \tilde{C} in a neighborhood U^0 of \mathbf{q}^0 sets the present approach to mechanical system kinematics and dynamics apart from the Maggi/Kane formulation, as presented in [5]. It is also distinct from a substantial literature [6–9] that derives ODE based on index 1 DAE that do not account for configuration or velocity constraints. The parametrization of Eq. (13) avoids approximation errors at the configuration level.

To see geometrically what Eq. (13) means and why the extent of the neighborhood V^0 is important, consider the situation schematically depicted in **Figure 2**. The vector $\mathbf{q}^0 + \mathbf{V}\mathbf{v}$ may be viewed as movement in the *tangent space* of \tilde{C} at \mathbf{q}^0 , which must be modified by vector $-\mathbf{U}\mathbf{u} = -\mathbf{U}\mathbf{h}(\mathbf{v})$ to yield $\mathbf{q}(\mathbf{v})$ on \tilde{C} . Eq. (13) consolidates the vectors into the continuously differentiable function $\mathbf{q} = \Psi(\mathbf{v})$. It is clear from **Figure 2**, however, that if \mathbf{v} is large, the vector $-\mathbf{U}\mathbf{u}$ that is perpendicular to the tangent space to \tilde{C} at \mathbf{q}^0 may not intersect \tilde{C} , in which case the parametrization fails;

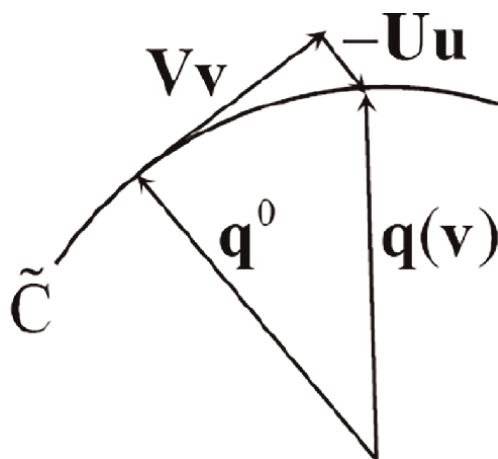


Figure 2.
 Projection onto \tilde{C} .

i.e., \mathbf{v} is no longer a valid vector of independent generalized coordinates. Fortunately, a large *condition number* of $\Phi_{\mathbf{q}}(\mathbf{q})\mathbf{U}$ [4] is an effective warning that such a situation is eminent, in which case a new vector $\bar{\mathbf{q}}^0$ is defined as the current value of \mathbf{q} , new bases $\bar{\mathbf{V}}$ and $\bar{\mathbf{U}}$ are evaluated, and analysis is continued with a new mapping $\bar{\Psi}(\bar{\mathbf{v}})$ on neighborhood \bar{U}^0 . The condition number of the kinematic matrix $\Phi_{\mathbf{q}}(\mathbf{q})\mathbf{U}$ provides the criterion for definition of charts in the atlas that defines the differentiable manifold \tilde{C} . This is a critical link between the kinematics of the system and its differentiable manifold.

The mapping of Eq. (13) is a *parameterization* of open subsets $U^i \subset \tilde{C}$ of the constraint manifold, equivalently the regular kinematic configuration space, by the variable \mathbf{v} on open subsets $V^i \subset \mathbb{R}^{ngc-nhc}$. This provides a local mathematical representation of the kinematics of a mechanical system on the linear space $\mathbb{R}^{ngc-nhc}$, establishing the fact that the system has $ngc - nhc$ degrees of freedom in V^0 . The essential nonlinearities of the kinematic system are embedded in the function $\Psi(\mathbf{v})$.

It is important to note that the set \tilde{C} of normalized Euler parameters used to define orientation of a body in space cannot be parameterized by a single mapping from \mathbb{R}^3 [10]. Thus, constraint manifolds of mechanical systems in space can generally not be parameterized by a single chart. This contradicts some mechanical system dynamics literature that assumes systems can be globally represented using independent generalized coordinates, called *Lagrangian coordinates* [11].

3.2 Computation of $\mathbf{h}(\mathbf{v})$

While the vector function $\mathbf{h}(\mathbf{v})$ of Eq. (13) is shown to exist and be a differentiable function of \mathbf{v} , the derivation does not show how it may be evaluated. Since it is central to implementing the parameterization defined by Eq. (13), numerical methods for its evaluation are needed. This is pivotal in computational kinematics and dynamics of mechanical systems. If an abstract mathematical approach to manifold theory had been employed, no insight into computation would be obtained. A critical component

in iterative computation of $\mathbf{u} = \mathbf{h}(\mathbf{v})$ is the matrix $\Phi_{\mathbf{q}}(\mathbf{q})\mathbf{U}$ that is nonsingular for $\mathbf{q} \in \tilde{\mathcal{C}}$. It defines the matrix

$$\mathbf{B}(\mathbf{q}) = (\Phi_{\mathbf{q}}(\mathbf{q})\mathbf{U})^{-1} \quad (14)$$

that will play a key role in evaluating $\mathbf{h}(\mathbf{v})$ and in obtaining ODE of mechanical system dynamics.

Computation in mechanical system dynamics is carried out on a time grid t^j , where t^0 is given and $t^j = t^0 + jh$, $j = 1, 2, \dots$, where $h > 0$ is a small step size. At t^0 , $\mathbf{u}^0 = \mathbf{0}$ and $\mathbf{B}^0 = (\mathbf{U}^T \mathbf{U})^{-1}$ is numerically evaluated. At t^j , presume $\mathbf{u}^j = \mathbf{h}(\mathbf{v}^j)$ and $\mathbf{B}^j = \mathbf{B}(\mathbf{q}^j)$ are known. Equations of motion of Section 6 are formed at t^j and integrated to obtain \mathbf{v}^{j+1} and \mathbf{q}^{j+1} . To solve Eq. (10) for $\mathbf{u}^{j+1} = \mathbf{h}(\mathbf{v}^{j+1})$, with a given \mathbf{v}^{j+1} , set $\mathbf{u}_1^{j+1} = \mathbf{u}^j$, where subscript is iteration number. Newton-Raphson iteration [12] for \mathbf{u}^{j+1} is

$$\Phi_{\mathbf{u}}^{j+1} \Delta \mathbf{u}_i^{j+1} = -\Phi_{\mathbf{q}}^{j+1} \mathbf{U} \Delta \mathbf{u}_i^{j+1} = -\Phi(\mathbf{q}^0 + \mathbf{V} \mathbf{v}^{j+1} - \mathbf{U} \mathbf{u}_i^{j+1})$$

Since $(\Phi_{\mathbf{q}}^{j+1} \mathbf{U})^{-1} \approx \mathbf{B}^j$ and Newton-Raphson iteration does not require an exact inverse [12], an approximate solution is $\Delta \mathbf{u}_i^{j+1} = \mathbf{B}^j \Phi(\mathbf{q}^0 + \mathbf{V} \mathbf{v}^{j+1} - \mathbf{U} \mathbf{u}_i^{j+1})$ and $\mathbf{u}_{i+1}^{j+1} = \mathbf{u}_i^{j+1} + \Delta \mathbf{u}_i^{j+1}$. This yields the iterative algorithm

$$\mathbf{u}^{i+1} = \mathbf{u}^i + \mathbf{B} \Phi(\mathbf{q}^0 + \mathbf{V} \mathbf{v} - \mathbf{U} \mathbf{u}^i), \quad i = 1, 2, \dots, \quad \text{until } \|\Phi(\mathbf{q}^0 + \mathbf{V} \mathbf{v}^{j+1} - \mathbf{U} \mathbf{u}_{i+1}^{j+1})\| \leq \text{utol} \quad (15)$$

where utol is a specified error tolerance. Since the Newton-Raphson method does not require an exact Jacobian, the matrix \mathbf{B}^j is held constant throughout the process. This is an efficient computation, requiring only matrix multiplication.

The matrix $\mathbf{B}(\mathbf{q}^{j+1})$ must satisfy Eq. (14), written in the form $\bar{\mathbf{R}} = (\Phi_{\mathbf{q}}(\mathbf{q})\mathbf{U})\mathbf{B}(\mathbf{q}) - \mathbf{I} = \mathbf{0}$. With an approximation $\mathbf{B}_1^{j+1} = \mathbf{B}(\mathbf{q}^j)$ of the solution and suppressing the argument \mathbf{q} , since it does not change in the iterative process for $\mathbf{B}(\mathbf{q})$, the matrix version of *Newton-Raphson iteration* [3, 12] is defined by $(\Phi_{\mathbf{q}}^{j+1} \mathbf{U}) \Delta \mathbf{B}_i^{j+1} = -\bar{\mathbf{R}}_i^{j+1} = -\Phi_{\mathbf{q}}^{j+1} \mathbf{U} \mathbf{B}_i^{j+1} + \mathbf{I}$. Since the matrix $(\Phi_{\mathbf{q}}^{j+1} \mathbf{U})$ need not be inverted with great precision for use in the Newton-Raphson process and $\mathbf{B}_i^{j+1} \approx (\Phi_{\mathbf{q}}^{j+1} \mathbf{U})^{-1}$, $\Delta \mathbf{B}_i^{j+1} = -\mathbf{B}_i^{j+1} \Phi_{\mathbf{q}}^{j+1} \mathbf{U} \mathbf{B}_i^{j+1} + \mathbf{B}_i^{j+1}$ and $\mathbf{B}^{i+1} = \mathbf{B}^i + \Delta \mathbf{B}^i$. This yields the iterative algorithm

$$\mathbf{B}_{i+1}^{j+1} = 2\mathbf{B}_i^{j+1} - \mathbf{B}_i^{j+1} \Phi_{\mathbf{q}}^{j+1} \mathbf{U} \mathbf{B}_i^{j+1}, \quad i = 1, 2, \dots, \quad \text{until } \|\Phi_{\mathbf{q}}^{j+1} \mathbf{U} \mathbf{B}_{i+1}^{j+1} - \mathbf{I}\| \leq \text{Btol} \quad (16)$$

where Btol is a specified error tolerance. This is an efficient computation, requiring only matrix multiplication. If the condition number of the matrix $\Phi_{\mathbf{q}}^{j+1} \mathbf{U}$ exceeds a tolerance that precludes accurate evaluation of its inverse \mathbf{B}^{j+1} the numerical process on the time grid is halted, a reparameterization is evaluated, and the process is continued on a new chart.

These results provide practical tools for *local parameterization* of the regular configuration manifold $\tilde{\mathcal{C}}$ and establish a fundamental linkage between topology and the engineering kinematics and dynamics of mechanical systems.

4. Global properties of constraint manifolds in Euclidean space

Results presented in Section 3 are local, in the sense that they are assured to hold only in neighborhoods of configurations involved. For example, a manifold \tilde{C} looks like the Euclidean space $\mathbb{R}^{ngc-nhc}$ on a chart (U, φ) , where U is an open subset of \tilde{C} and $\varphi : U \rightarrow \mathbb{R}^{ngc-nhc}$ is a homeomorphism onto some open subset of $\mathbb{R}^{ngc-nhc}$. A metaphor is that a near-sighted person sees the neighborhood U as a Euclidean topological space, even though \tilde{C} in its entirety is not a Euclidean space. This is akin to the early belief that the earth is flat. A powerful impact of differential geometry is that it extends local results to global forms; akin to an optometrist providing corrective lenses to the near-sighted person that sees the earth as flat, showing its nonlinear character by seeing well beyond a neighborhood of one's feet. Differential geometry is thus a tool, analogous to corrective lenses, that enables observation in the large.

Our local parametrization of \tilde{C} is based on the Implicit Function Theorem, which implies that the intrinsic topology on \tilde{C} given by the differentiable atlas coincides with the topology that \tilde{C} inherits as a subspace of \mathbb{R}^{ngc} (see Section 2). In fact, \tilde{C} is a differentiable $(ngc - nhc)$ -dimensional submanifold of \mathbb{R}^{ngc} . In other words, the number of degrees of freedom is constant (and equal to $(ngc - nhc)$) throughout the regular configuration spaces. This is not the case on the entire configuration space C , which is the principal reason why we restrict our attention to \tilde{C} .

In general, \tilde{C} is not a path-connected topological space, even if C itself is path-connected, see for example the Triple Slider-Crank Mechanism in Section 2.4. The path-components of \tilde{C} are maximal, path connected, open subsets of \tilde{C} , in each of which the mechanism is locally parameterized by Eq. (13). From a kinematic viewpoint, these path-components are maximal singularity free domains of kinematic functionality, abbreviated *domains of functionality*.

Extension of the local parameterization of Eq. (13) to a global setting on each \tilde{C}_α is a powerful result that differential geometry brings to kinematics. While such results are consequences of differential geometry, a deep study of the abstract theory of differential geometry is not required for their use. Once established, implementation of these results requires only use of linear algebra and multivariable calculus.

Given two regular configurations c_1, c_2 in the configuration space C several situations may arise. The two configurations may belong to distinct path-components of C so that there does not exist a continuous change of the states of the mechanism that would transform c_1 into c_2 . In general this means that to achieve a transition from the configuration c_1 to the configuration c_2 , the mechanism must be disassembled and then reassembled in some different way. It can also happen that c_1 and c_2 belong to the same path-component of C but are in different domains of functionality, say \tilde{C}_1 and \tilde{C}_2 , of the regular configuration space \tilde{C} . Different path-components of a manifold are not only disjoint but are what is called *topologically separated* in the sense that the closure of \tilde{C}_1 does not intersect \tilde{C}_2 and vice-versa. This means that every trajectory in the configuration space C along which the mechanism is transformed from c_1 to c_2 must cross the frontiers of the respective domains of functionality and pass through at least one singularity. In practice this means that in order to pass from c_1 to c_2 , external influence may be exerted at some points to make a continuous transition between different domains of functionality, but the mechanism does not need to be disassembled and reassembled in the process. Nevertheless, it encounters a singularity.

In general, a path between configurations that belong to the same domain of functionality may be constructed as indicated below. If two charts (U_i, φ_i) and (U_j, φ_j) defined on \tilde{C} overlap, as shown schematically in **Figure 3**, then there is a differentiable mapping $\varphi_j \varphi_i^{-1} : \varphi_i(U_i \cap U_j) \rightarrow \varphi_j(U_i \cap U_j)$ between intersecting charts and the charts are called *compatible*, much as two planar maps of the surface of the earth that cover the same locale are compatible. As explained in Section 2.2 a family of *compatible charts* whose images cover the entire manifold is called an *atlas*. The atlas provides a structure upon which to characterize trajectories on the constraint manifold, such as a solution curve for the equations of constrained dynamics, as shown schematically in **Figure 4**. The charts provide for changes in local coordinates along a trajectory $\mathbf{q}(\mathbf{v}(t))$ in \tilde{C}_α , including differentiability of generalized coordinates and tangents. The theory of differentiable manifolds [2] assures that a solution trajectory can be continued in the open submanifold \tilde{C} until the constraint Jacobian $\Phi_{\mathbf{q}}(\mathbf{q})$ becomes rank deficient. While the theory does not explicitly determine the extent of

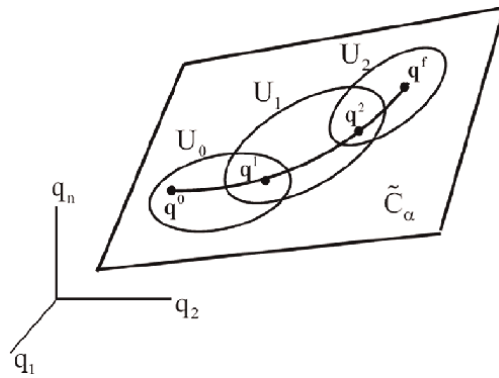


Figure 3.
Continuation of solution trajectory over charts.

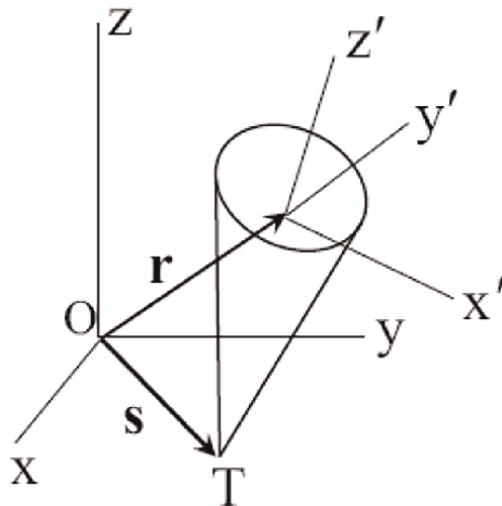


Figure 4.
Symmetric top with tip constrained to be unit distance from origin O .

components of the regular configuration manifold, it does assure that the parameterization by charts that has been presented is valid on the largest possible subset of regular configuration space. Nothing more is possible.

5. Kinematics as a function of time

In Section 3 we obtained an explicit expression (Eq. 13) for configurations $\mathbf{q} \in \tilde{C}$ in terms of local coordinates $\mathbf{v} \in \mathbb{R}^{ngc-nhc}$. In this section we continue this work and give explicit expressions for $\dot{\mathbf{q}}$ and $\ddot{\mathbf{q}}$ in terms of \mathbf{v} , $\dot{\mathbf{v}}$ and $\ddot{\mathbf{v}}$. For dynamics of mechanical systems, generalized coordinates are a function of time, $\mathbf{q} = \mathbf{q}(t)$. By differentiating Eq. (1) with respect to time we obtain a *velocity constraint equation*

$$\Phi_{\mathbf{q}}(\mathbf{q})\dot{\mathbf{q}} = \mathbf{0} \quad (17)$$

Furthermore, a second derivative with respect to time yields $(\Phi_{\mathbf{q}}(\dot{\mathbf{q}}))\dot{\mathbf{q}} + \Phi_{\mathbf{q}}(\mathbf{q})\ddot{\mathbf{q}} = \mathbf{0}$ (where $(\Phi_{\mathbf{q}}(\dot{\mathbf{q}}))$ denotes the component-wise derivative of the Jacobian matrix $\Phi_{\mathbf{q}}(\mathbf{q})$ with respect to t). Thus we obtain an *acceleration constraint equation*

$$\Phi_{\mathbf{q}}(\mathbf{q})\ddot{\mathbf{q}} = -(\Phi_{\mathbf{q}}(\dot{\mathbf{q}}))\dot{\mathbf{q}} = -\gamma(\mathbf{q}, \dot{\mathbf{q}}), \quad (18)$$

where we abbreviated

$$\gamma(\mathbf{q}, \dot{\mathbf{q}}) := (\Phi_{\mathbf{q}}(\dot{\mathbf{q}}))\dot{\mathbf{q}} \quad (19)$$

Eqs. (1), (17) and (18) represent three forms of constraint that must be satisfied by $\mathbf{q} \in \tilde{C}$.

With time held fixed, Eq. (17) yields

$$\Phi_{\mathbf{q}}(\mathbf{q})\delta\mathbf{q} = \mathbf{0} \quad (20)$$

which defines kinematically admissible *virtual displacements* $\delta\mathbf{q}$ [3, 11] that are tangent to the regular configuration space.

Let us begin with two auxiliary results that will be used in computation. By differentiating Eq. (13) with respect to t , we get

$$\dot{\mathbf{q}} = \mathbf{V}\dot{\mathbf{v}} - \mathbf{U}\dot{\mathbf{h}}, \quad (21)$$

which together with Eq. (17) yields

$$\Phi_{\mathbf{q}}(\mathbf{q})\mathbf{V}\dot{\mathbf{v}} = \Phi_{\mathbf{q}}(\mathbf{q})\mathbf{U}\dot{\mathbf{h}} \quad (22)$$

Furthermore, in Section 3 we observed that $\Phi_{\mathbf{q}}(\mathbf{q})\mathbf{U}$ is invertible in a neighborhood of \mathbf{q}^0 , so we may define the matrix inverse

$$\mathbf{B}(\mathbf{q}) := (\Phi_{\mathbf{q}}(\mathbf{q})\mathbf{U})^{-1}.$$

of Eq. (14). By differentiating the equality $\mathbf{B}\Phi_{\mathbf{q}}\mathbf{U} = \mathbf{I}$ with respect to t we obtain $\dot{\mathbf{B}}\Phi_{\mathbf{q}}\mathbf{U} + \mathbf{B}\dot{\Phi}_{\mathbf{q}}\mathbf{U} = \mathbf{0}$, therefore

$$\dot{\mathbf{B}}\Phi_{\mathbf{q}}U = -\mathbf{B}\dot{\Phi}_{\mathbf{q}}U \quad (23)$$

It is now easy to compute $\dot{\mathbf{q}}$ (we suppress arguments):

$$\begin{aligned} \dot{\mathbf{q}} &= \mathbf{V}\dot{\mathbf{v}} - \mathbf{U}\dot{\mathbf{h}} && \text{by Eq. (21)} \\ &= \mathbf{V}\dot{\mathbf{v}} - \mathbf{U}\mathbf{B}\Phi_{\mathbf{q}}\mathbf{U}\dot{\mathbf{h}} && \text{using } \mathbf{B}\Phi_{\mathbf{q}}\mathbf{U} = \mathbf{I} \\ &= \mathbf{V}\dot{\mathbf{v}} - \mathbf{U}\mathbf{B}\Phi_{\mathbf{q}}\mathbf{V}\dot{\mathbf{v}} && \text{by Eq. (22)} \\ &= \mathbf{D}\dot{\mathbf{v}} && \text{we abbreviate } \mathbf{D}(\mathbf{q}) := (\mathbf{I} - \mathbf{U}\mathbf{B}(\mathbf{q})\Phi_{\mathbf{q}}(\mathbf{q}))\mathbf{V} \end{aligned} \quad (24)$$

Observe that the matrix function $\mathbf{D}(\mathbf{q})$ satisfies

$$\Phi_{\mathbf{q}}\mathbf{D} = (\Phi_{\mathbf{q}} - \Phi_{\mathbf{q}}\mathbf{U}\mathbf{B}\Phi_{\mathbf{q}})\mathbf{V} = (\Phi_{\mathbf{q}} - \Phi_{\mathbf{q}})\mathbf{V} = \mathbf{0}, \quad (25)$$

which means that the columns of $\mathbf{D}(\mathbf{q})$ comprise a continuously differentiable basis for the null space of $\Phi_{\mathbf{q}}$.

In order to express $\ddot{\mathbf{q}}$, we differentiate Eq. (24) with respect to t :

$$\ddot{\mathbf{q}} = \mathbf{D}\ddot{\mathbf{v}} + \dot{\mathbf{D}}\dot{\mathbf{v}} \quad (26)$$

and compute

$$\begin{aligned} \dot{\mathbf{D}}\dot{\mathbf{v}} &= -(\mathbf{U}\dot{\mathbf{B}}\Phi_{\mathbf{q}}\mathbf{V} + \mathbf{U}\mathbf{B}\dot{\Phi}_{\mathbf{q}}\mathbf{V})\dot{\mathbf{v}} \\ &= -\mathbf{U}\dot{\mathbf{B}}\Phi_{\mathbf{q}}\mathbf{U}\dot{\mathbf{h}} - \mathbf{U}\mathbf{B}\dot{\Phi}_{\mathbf{q}}\mathbf{V}\dot{\mathbf{v}} && \text{by Eq. (22)} \\ &= \mathbf{U}\mathbf{B}\dot{\Phi}_{\mathbf{q}}\mathbf{U}\dot{\mathbf{h}} - \mathbf{U}\mathbf{B}\dot{\Phi}_{\mathbf{q}}\mathbf{V}\dot{\mathbf{v}} && \text{by Eq. (23)} \\ &= \mathbf{U}\mathbf{B}\dot{\Phi}_{\mathbf{q}}(\mathbf{U}\dot{\mathbf{h}} - \mathbf{V}\dot{\mathbf{v}}) = -\mathbf{U}\mathbf{B}\gamma && \text{by Eq. (19)} \end{aligned} \quad (27)$$

By combining Eqs. (26) and (27) we obtain

$$\ddot{\mathbf{q}} = \mathbf{D}(\mathbf{q})\ddot{\mathbf{v}} - \mathbf{U}\mathbf{B}(\mathbf{q})\gamma(\mathbf{q}, \dot{\mathbf{q}}) \quad (28)$$

We may finally summarize Eqs. (13), (24) and (28) that define $\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}}$ as functions of $\mathbf{v}, \dot{\mathbf{v}}, \ddot{\mathbf{v}}$.⁴

$$\begin{aligned} \mathbf{q}(\mathbf{v}) &= \mathbf{q}^0 + \mathbf{V}\mathbf{v} - \mathbf{U}\mathbf{h}(\mathbf{v}) \\ \dot{\mathbf{q}}(\mathbf{v}, \dot{\mathbf{v}}) &= \mathbf{D}(\mathbf{q}(\mathbf{v}))\dot{\mathbf{v}} \\ \ddot{\mathbf{q}}(\mathbf{v}, \dot{\mathbf{v}}, \ddot{\mathbf{v}}) &= \mathbf{D}(\mathbf{q}(\mathbf{v}))\ddot{\mathbf{v}} - \mathbf{U}\mathbf{B}(\mathbf{q}(\mathbf{v}))\gamma(\mathbf{q}(\mathbf{v}), \dot{\mathbf{q}}(\mathbf{v}, \dot{\mathbf{v}})) \end{aligned} \quad (29)$$

We have thus demonstrated that $\mathbf{q}, \dot{\mathbf{q}}$ and $\ddot{\mathbf{q}}$ are differentiable functions of $\mathbf{v}, \dot{\mathbf{v}}$ and $\ddot{\mathbf{v}}$ and provided explicit formulae to describe $\mathbf{q}, \dot{\mathbf{q}}$ and $\ddot{\mathbf{q}}$ that satisfy constraint Eqs. (1), (17) and (18), once independent coordinates $\mathbf{v}, \dot{\mathbf{v}}$ and $\ddot{\mathbf{v}}$ are determined as a solution of equations of dynamics presented in Section 6.

The importance of the fact that generalized coordinates \mathbf{q} and their derivatives $\dot{\mathbf{q}}$ and $\ddot{\mathbf{q}}$ identically satisfy the holonomic, velocity and acceleration constraints in a neighborhood of \mathbf{q}^0 cannot be overemphasized. This aspect of the differentiable manifold formulation is critical in avoiding so called *constraint drift* that is associated with many mechanical system dynamics formulations that are based on approximations and their numerical implementations with projections onto the constraint manifold, see [6–9, 13, 14].

6. Manifold-based ODE of mechanical system dynamics

Using the foregoing manifold-based kinematics formulation, the *variational equation of motion* for a holonomic system is reduced to a system of ODE. The d'Alembert variational equation of motion for a mechanical system is [3, 11].

$$\delta \mathbf{q}^T (\mathbf{M} \mathbf{q} \ddot{\mathbf{q}} - \mathbf{S}(\mathbf{q}, \dot{\mathbf{q}}) - \mathbf{Q}^A(\mathbf{q}, \dot{\mathbf{q}}, t)) = 0 \quad (30)$$

which must hold for all $\delta \mathbf{q}$ that satisfy Eq. (20); i.e., $\Phi_{\mathbf{q}} \delta \mathbf{q} = \mathbf{0}$, for a system with holonomic constraints of Eq. (1) that include Euler parameter normalization conditions for spatial bodies. Terms appearing in Eq. (30) are the mass matrix $\mathbf{M}(\mathbf{q})$ that is symmetric and positive definite on the null space of $\Phi_{\mathbf{q}}(\mathbf{q})$ [3], velocity coupling terms $\mathbf{S}(\mathbf{q}, \dot{\mathbf{q}})$ that are quadratic in $\dot{\mathbf{q}}$, and generalized applied force $\mathbf{Q}^A(\mathbf{q}, \dot{\mathbf{q}}, t)$. It should be noted that for spatial systems with orientation defined by Euler parameters, $\mathbf{M}(\mathbf{q})$ is singular. Nevertheless, it is positive definite on the null space of $\Phi_{\mathbf{q}}(\mathbf{q})$.

As in Eq. (24), with time held fixed, $\delta \mathbf{q} = \mathbf{D}(\mathbf{q}) \delta \mathbf{v}$. Using this relation and Eq. (29), Eq. (30) reduces to a variational equation in \mathbf{v} ,

$$\delta \mathbf{v}^T \mathbf{D}^T(\mathbf{q}, t) (\mathbf{M}(\mathbf{q}) \mathbf{D}(\mathbf{q}) \ddot{\mathbf{v}} - \mathbf{M}(\mathbf{q}) \mathbf{U} \mathbf{B}(\mathbf{q}) \gamma(\mathbf{q}, \dot{\mathbf{q}}) - \mathbf{S}(\mathbf{q}, \dot{\mathbf{q}}) - \mathbf{Q}^A(\mathbf{q}, \dot{\mathbf{q}}, t)) = 0 \quad (31)$$

From Eq. (25),

$$\Phi_{\mathbf{q}}(\mathbf{q}) \delta \mathbf{q} = \Phi_{\mathbf{q}}(\mathbf{q}) \mathbf{D}(\mathbf{q}) \delta \mathbf{v} = \mathbf{0}$$

for arbitrary $\delta \mathbf{v}$. Thus, Eq. (31) holds for arbitrary $\delta \mathbf{v}$ and, suppressing arguments,

$$\mathbf{D}^T \mathbf{M} \mathbf{D} \ddot{\mathbf{v}} = \mathbf{D}^T (\mathbf{M} \mathbf{U} \mathbf{B} \gamma + \mathbf{S} + \mathbf{Q}^A) \quad (32)$$

where Eq. (29) shows that \mathbf{q} and $\dot{\mathbf{q}}$ are functions of \mathbf{v} and $\dot{\mathbf{v}}$, and

$$\begin{aligned} \mathbf{M}(\mathbf{q}(\mathbf{v})) &= \mathbf{M}(\mathbf{v}) \\ \gamma(\mathbf{q}(\mathbf{v}), \dot{\mathbf{q}}(\mathbf{v}, \dot{\mathbf{v}})) &= \gamma(\mathbf{v}, \dot{\mathbf{v}}) \\ \mathbf{S}(\mathbf{q}(\mathbf{v}), \dot{\mathbf{q}}(\mathbf{v}, \dot{\mathbf{v}})) &= \mathbf{S}(\mathbf{v}, \dot{\mathbf{v}}) \\ \mathbf{Q}^A(\mathbf{q}(\mathbf{v}), \dot{\mathbf{q}}(\mathbf{v}, \dot{\mathbf{v}}), t) &= \mathbf{Q}^A(\mathbf{v}, \dot{\mathbf{v}}, t) \\ \mathbf{B}(\mathbf{q}(\mathbf{v})) &= \mathbf{B}(\mathbf{v}) \\ \mathbf{D}(\mathbf{q}(\mathbf{v})) &= \mathbf{D}(\mathbf{v}) \end{aligned} \quad (33)$$

Eq. (32) is thus a second order ODE in \mathbf{v} .

At an initial configuration \mathbf{q}^0 that satisfies Eq. (1), $\Phi_{\mathbf{q}}(\mathbf{q}^0) \mathbf{V} = \mathbf{0}$ and $\mathbf{D}^0 = \mathbf{D}(\mathbf{q}^0) = \mathbf{V} - \mathbf{U} \mathbf{B}(\mathbf{q}^0) \Phi_{\mathbf{q}}(\mathbf{q}^0) \mathbf{V} = \mathbf{V}$. Thus, $\mathbf{D}^{0T} \mathbf{M}(\mathbf{q}^0) \mathbf{D}^0 = \mathbf{V}^T \mathbf{M}(\mathbf{q}^0) \mathbf{V}$. Since the columns of \mathbf{V} form a basis for the null-space of $\Phi_{\mathbf{q}}(\mathbf{q}^0)$, on which $\mathbf{M}(\mathbf{q}^0)$ is positive definite, the symmetric reduced mass matrix $\mathbf{D}^T(\mathbf{v}) \mathbf{M}(\mathbf{v}) \mathbf{D}(\mathbf{v})$ of Eq. (32), which is a continuous function of \mathbf{v} , is positive definite, hence nonsingular, in a neighborhood of $\mathbf{v}^0 = \mathbf{0}$. Multiplying both sides of Eq. (32) by the inverse of $\mathbf{D}^T \mathbf{M} \mathbf{D}$, suppressing arguments of functions involved for notational simplicity, yields the traditional form of a second order ODE,

$$\ddot{\mathbf{v}} = (\mathbf{D}^T \mathbf{M} \mathbf{D})^{-1} \mathbf{D}^T (\mathbf{M} \mathbf{U} \mathbf{B} \gamma + \mathbf{S} + \mathbf{Q}^A) := \mathbf{g}(\mathbf{v}, \dot{\mathbf{v}}, t) \quad (34)$$

Given initial values of $\mathbf{q}(t^0)$ and $\dot{\mathbf{q}}(t^0)$ that satisfy Eqs. (1) and (17) at t^0 we obtain initial conditions on \mathbf{v} and $\dot{\mathbf{v}}$,

$$\begin{aligned}\mathbf{v}(t^0) &= \mathbf{0} \\ \dot{\mathbf{v}}(t^0) &= \mathbf{V}^T \dot{\mathbf{q}}(t^0)\end{aligned}\tag{35}$$

Assuming all functions that appear in Eq. (33) are continuously differentiable with respect to their arguments, the initial-value problem of Eqs. (34) and (35) has a unique solution for $\mathbf{v}(t)$ in a neighborhood of (\mathbf{v}^0, t^0) [15]. As shown in Section 4.1, this solution can be continued on the constraint manifold shown schematically in **Figure 3** until a singular configuration or the desired final time t^f is encountered. Further, if terms in Eq. (33) are k -times continuously differentiable with respect to a vector \mathbf{b} of parameters, so is the solution [15]. With these existence, uniqueness, and differentiability properties, the initial-value problem of Eqs. (34) and (35) is well posed.

Numerical reduction of variational equations of motion to ODE depends on tolerances $utol$ in approximating \mathbf{u} and $Btol$ in approximating \mathbf{B} in Eqs. (15) and (16). Thus, after numerical solution of the ODE of motion for \mathbf{v} and its derivatives, Eq. (29) yields $\mathbf{q}, \dot{\mathbf{q}}, \ddot{\mathbf{q}}$ that satisfy the constraints of Eqs. (1), (17), and (18), to within specified error tolerances. More specifically, the accuracy with which the configuration generalized coordinates \mathbf{q} are evaluated depends on the accuracy with which \mathbf{u} is computed in Eq. (15); i.e., $utol$. As long as convergence in this iteration is achieved, the accuracy with which \mathbf{B} is computed in Eq. (15) is immaterial. In contrast, the accuracy of $\dot{\mathbf{q}}$ and $\ddot{\mathbf{q}}$ that are computed in Eq. (29) is directly impacted by the accuracy with which \mathbf{B} is computed in Eq. (16); i.e., $Btol$. Thus, both $utol$ and $Btol$ control the accuracy with which velocity and acceleration generalized coordinates are evaluated.

It is noted that with the unique solution of Eqs. (34) and (35) and its derivatives known, Eq. (29) determines the unique solution $\mathbf{q}(t)$ and its derivatives of the mechanical system dynamics problem. As emphasized earlier, this solution satisfies all three forms of holonomic constraint, to within specified tolerances. Thus, problems associated with drift from constraints that arise in numerical treatments of the DAE of mechanical system dynamics that fail to enforce all forms of constraint are not of concern. Similarly, errors induced by use of an approximate ODE and modification of results to satisfy constraints are avoided.

7. Numerical examples

Two examples are treated using the manifold-based ODE formulation. The first is a five degree of freedom spatial spinning Top whose tip is constrained to be a unit distance from a point fixed in space. The second example is the planar triple slider-crank of Section 2.4. Its dynamics in two of the four disjoint singularity free domains of functionality defined in Eq. (7) are simulated using an extended four body planar model with generalized coordinates in $\mathbb{R}^{ngc} = \mathbb{R}^{12}$, while retaining its two degrees of freedom.

7.1 Symmetric top with tip constrained to be a unit distance from a fixed point

Consider a symmetric Top with its tip that is 1 m from the centroid constrained to be 1 m from the origin of the $x - y - z$ frame, as shown in **Figure 4**. In this model,

$\mathbf{q} = [\mathbf{r}^T \ \mathbf{p}^T]^T$, where $\mathbf{p} \in \mathbb{R}^4$ is the vector of Euler parameters defined in the Appendix that characterizes orientation of the body fixed $x' - y' - z'$ reference frame. In addition to the Euler parameter normalization constraint of Eq. (37) in the Appendix, the length of vector $\mathbf{s} = \mathbf{r} - \mathbf{A}(\mathbf{p})\mathbf{u}_{z'}$ shown in **Figure 4** must be unity, i.e.,

$$\mathbf{s}^T \mathbf{s} = (\mathbf{r} - \mathbf{A}(\mathbf{p})\mathbf{u}_{z'})^T (\mathbf{r} - \mathbf{A}(\mathbf{p})\mathbf{u}_{z'}) = \mathbf{r}^T \mathbf{r} - 2\mathbf{r}^T \mathbf{A}(\mathbf{p})\mathbf{u}_{z'} + 1 = 1$$

Holonomic constraints for the system are thus

$$\Phi(\mathbf{q}) = \frac{1}{2} \begin{bmatrix} \mathbf{r}^T \mathbf{r} - 2\mathbf{r}^T \mathbf{A}(\mathbf{p})\mathbf{u}_{z'} \\ \mathbf{p}^T \mathbf{p} - 1 \end{bmatrix} = \mathbf{0}$$

Using Eq. (39) of the Appendix, the constraint Jacobian is

$$\Phi_{\mathbf{q}}(\mathbf{q}) = \begin{bmatrix} \mathbf{r}^T - \mathbf{u}_{z'}^T \mathbf{A}^T(\mathbf{p}) & -\mathbf{r}^T \mathbf{B}(\mathbf{p}, \mathbf{u}_{z'}) \\ \mathbf{0} & \mathbf{p}^T \end{bmatrix}$$

To determine if the Jacobian has full rank, form the equation

$$\Phi_{\mathbf{q}}(\mathbf{q})^T \begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \begin{bmatrix} (\mathbf{r} - \mathbf{A}(\mathbf{p})\mathbf{u}_{z'})\alpha \\ -\mathbf{B}^T(\mathbf{p}, \mathbf{u}_{z'})\mathbf{r}\alpha + \mathbf{p}\beta \end{bmatrix} = \begin{bmatrix} \mathbf{s}\alpha \\ -\mathbf{B}^T(\mathbf{p}, \mathbf{u}_{z'})\mathbf{r}\alpha + \mathbf{p}\beta \end{bmatrix} = \mathbf{0}$$

so $\mathbf{s}\alpha = \mathbf{0}$ and, since \mathbf{s} is a unit vector, $\alpha = 0$. The second equation thus reduces to $\mathbf{p}\beta = \mathbf{0}$ and, since \mathbf{p} is a unit vector, $\beta = 0$. This shows that the transpose of the Jacobian has full column rank, so the Jacobian has full rank for all \mathbf{q} . This shows that, with the parameterization of Section 3.2, $\tilde{C} = C$ is a differentiable manifold and there are no singularities in the configuration space. However [10], \tilde{C} cannot be parameterized by a single chart, so the continuation process of Section 4.1 must be employed. In this example, $ngc = 7$ and $nhc = 2$, so there are five degrees of freedom, $\mathbf{B}(\mathbf{q})$ is a 2×2 matrix function, and $\mathbf{h}(\mathbf{v})$ is a two-vector function.

Inertia properties of the Top are $m = 30 \text{ kg}$ and $\mathbf{J}' = \text{diag}(90, 90, 30) \text{ kg} \cdot \text{m}^2$. The initial configuration of the system is with the tip on the z -axis one unit below the origin, the centroid at the origin, and the z' -axis coincident with the z -axis, so $\mathbf{q}^0 = [0 \ 0 \ 0 \ 1 \ 0 \ 0 \ 0]^T$. The initial velocity is $\dot{\mathbf{r}}^0 = \mathbf{0}$. The initial angular velocity in the body fixed $x' - y' - z'$ reference frame is $\omega'^0 = [\varepsilon \ \varepsilon \ \omega z^0]$, where initial angular velocity components $\varepsilon = 10^{-12}$ about the x' and y' axes play the role of perturbations from the vertical configuration. From Eq. (38) of the Appendix, $\mathbf{p}^0 = 0.5 \mathbf{G}(\mathbf{p}^0)^T \omega'^0$. As a check on *conservation of energy*, total energy is $TE = 2\dot{\mathbf{p}}^T \mathbf{G}(\mathbf{p})^T \mathbf{J}' \mathbf{G}(\mathbf{p}) \dot{\mathbf{p}} + mgr_z$, where r_z is the z component of the vector \mathbf{r} .

General-purpose MATLAB Code 5.9 of [3] that is available for download at no cost from ResearchGate has been used to implement the manifold-based ODE formulation for this example. Simulation results in **Figure 5** obtained using the fourth order Nystrom4 explicit integrator [16] with constant step size 0.001 s show the Top stabilizing at an initial angular velocity about the z' -axis of approximately 13.5 rad/s. The upper plots for each initial angular velocity are $x - y$ trajectories of the centroid and the lower plots are $x - y$ trajectories of the tip. Total energy variation for this

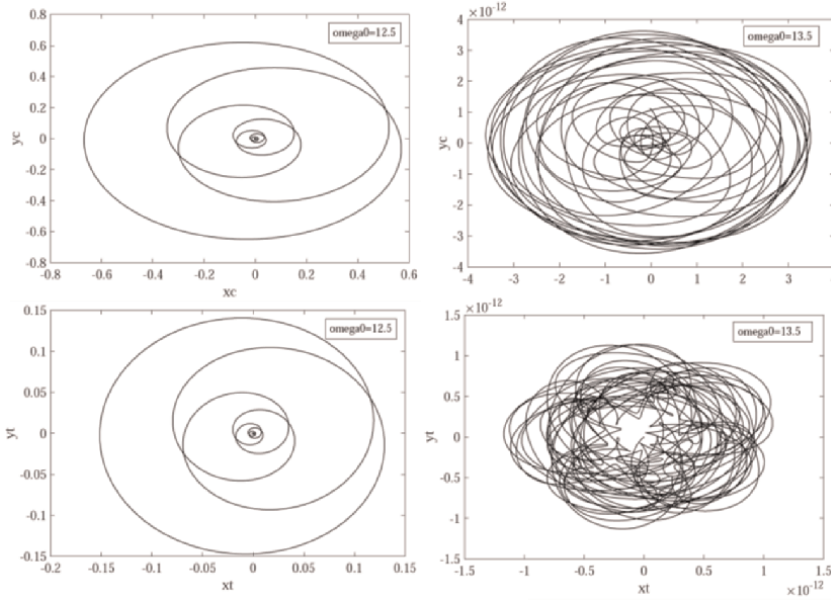


Figure 5.
Plots of centroid and tip x - y trajectories for top on Bar.

conservative system was $2e-4\%$ over the 100 s simulation. On average, charts were algorithmically reparametrized 524 times during e5 time steps in each simulation, all due to a limit of 0.75 on the norm of \mathbf{v} , or once per 191 time steps. Only 0.02% of CPU time and no person effort was devoted to reparameterization.

As evidence that the formulation provides accurate results, with $\text{Tol} = \text{Btol} = \text{utol}$, where Btol and utol are defined in Eqs. (15) and (16), the maximum norm of error in position, velocity, and acceleration constraints over 100 s simulations are reported in **Table 1**. As predicted by the theory, these errors approach zero to computer precision as Tol approaches zero.

7.2 Triple slider-crank

For numerical treatment of the triple slider-crank of **Figure 1**, the mechanism is modeled as four planar bodies, each with three generalized coordinates $\mathbf{q}_i = [r_i^T \ \varphi_i]^T$, so $\text{ngc} = 12$. Three translational constraints, one revolute constraint, and two distance constraints are imposed, for a total of $\text{nhc} = 10$ constraints. For simulation using the manifold-based ODE formulation, therefore, $\mathbf{B}(\mathbf{q})$ is a 10×10 matrix function of 12 generalized coordinates and $\mathbf{h}(\mathbf{v})$ is a 10-vector function of two independent variables. The dimensionality of the formulation is thus much higher than the spatial spinning Top. Simulation using the manifold-based ODE formulation was carried out using MATLAB Code 5.7 of [3].

In order to avoid singular configurations identified in Section 2.3, a torque τ shown in **Figure 1** is defined on \tilde{C}_1 and \tilde{C}_3 of Eq. (7) as

$$\tau = \frac{Kq_2}{(\pi/2 - |q_2|)^2}$$

Tol	Pos constr err	Vel constr err	Acc constr err
e-6	2e-8	e-6	5e-5
e-8	2e-10	3e-8	2e-7
e-10	2e-12	2e-10	4e-9

Table 1.
Maximum norm of position, velocity, and acceleration constraint error.

A similar torque will assure avoidance of singularities in \tilde{C}_2 and \tilde{C}_4 . For purposes of simulation, parameters are selected as $K = 100$ N/m; inertia properties in the mks system are $m_1 = 2, J_1 = 1, m_2 = 10, J_2 = 10, m_3 = 2, J_3 = 1, m_4 = 2$ and $J_4 = 1$; and initial conditions in \tilde{C}_3 are $q_1^0 = 0, q_2^0 = 0, q_3^0 = 2, q_4^0 = 2, \dot{q}_1^0 = -6, \dot{q}_2^0 = 2, \dot{q}_3^0 = -6, \dot{q}_4^0 = 0$. For simulation in \tilde{C}_1 , the same data are used, except with the initial condition $q_3^0 = 0$.

Simulation in \tilde{C}_3 is carried out using the fourth order Nystrom4 explicit integrator and a second order trapezoidal implicit integrator [3]. Identical results were obtained for each integrator and, since the ODE is not stiff, the Nystrom4 integrator was used to carry out multiple simulations. Plots of $q_i, i = 1, 2, 3, 4$ for the simulation in \tilde{C}_3 are shown in **Figure 6**. Results for simulation with initial condition $q_3^0 = 0$ in \tilde{C}_1 are shown in **Figure 7**. In both simulations, trajectories are confined to the component in which they were initiated, consistent with the theory.

In each simulation, charts were algorithmically reparametrized 6 times during 2000 time steps of simulation, or once per 333 time steps. Only 0.068% of CPU time and no person effort was devoted to reparameterization.

As evidence that the formulation provides accurate results, with Tol = Btol = utol, where Btol and utol are defined in Eqs. (15) and (16), the maximum norm of error in position, velocity, and acceleration constraints over three simulations in \tilde{C}_3 are reported in **Table 2**. As predicted by the theory, these errors approach zero to computer precision as Tol approaches zero.

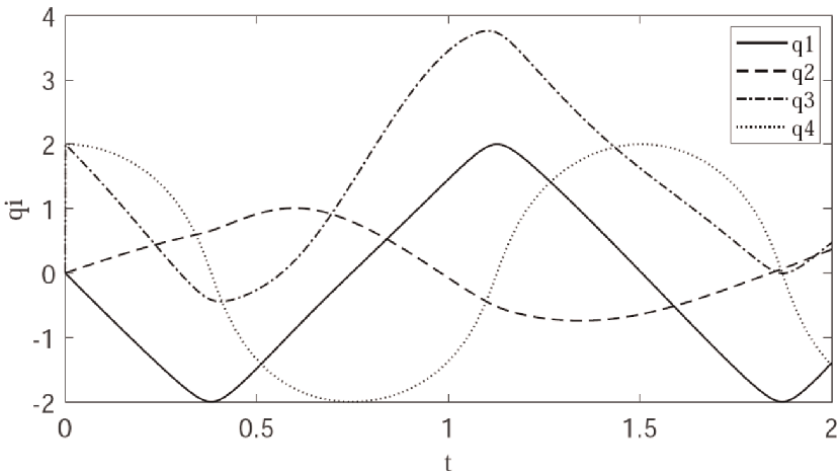


Figure 6.
Triple slider-crank simulation in \tilde{C}_3 .

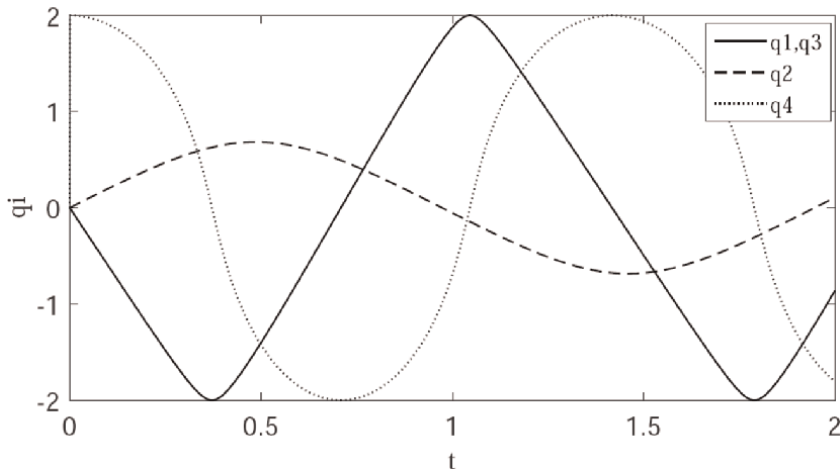


Figure 7.
Triple slider-crank simulation in \tilde{C}_1 .

Tol	Pos constr err	Vel constr err	Acc constr err
e-6	6e-12	9e-7	9e-6
e-8	5e-12	8e-9	8e-8
e-10	7e-15	7e-11	4e-10

Table 2.
Maximum norm of position, velocity, and acceleration constraint error.

8. Conclusions

Topological properties of Euclidean space that are summarized in Section 2 provide a rich foundation for differentiable manifold-based mechanical system kinematics and dynamics. Topological vector space attributes of Euclidean space are shown to be adequate for systematic treatment of mechanical system dynamics, without the need for more abstract theories of differential geometry that do not represent properties of mechanical system kinematics in Euclidean space. In the Euclidean setting, with modest results from multivariable calculus, local parameterization of the regular configuration space is carried out to constructively define a differentiable manifold structure that leads to Lagrange multiplier-free, well-posed ODE of mechanical system dynamics on maximal, singularity free, path connected components of regular configuration space.

Two examples demonstrate that error control is maintained for all three forms of holonomic kinematic constraint and that reparameterizations required in mechanical system dynamic simulation are algorithmically implemented, without person interaction, and require only approximately 0.02% of simulation CPU time. This suggests that alarm expressed in the engineering literature regarding the onerous overhead and complexity of reparameterization on a manifold is unwarranted. In fact, for spatial system simulation, since global orientation representation with any set of three parameters is impossible, numerous reparameterizations are unavoidable.

A. Euler parameter orientation generalized coordinates

With rotation of an orthogonal $x' - y' - z'$ frame by a counter-clockwise angle χ about a unit vector \mathbf{u} in the orthogonal $x - y - z$ frame, four Euler parameters $e_0 := \cos(\chi/2)$ and $\mathbf{e} := \sin(\chi/2)\mathbf{u}$ define the orthogonal orientation transformation matrix from the $x' - y' - z'$ frame to the $x - y - z$ frame as [3].

$$\mathbf{A} = (e_0^2 - \mathbf{e}^T \mathbf{e})\mathbf{I} + 2\mathbf{e}\mathbf{e}^T + 2e_0\tilde{\mathbf{e}} \quad (36)$$

where $\tilde{\mathbf{e}} = \begin{bmatrix} 0 & -e_3 & e_2 \\ e_3 & 0 & -e_1 \\ e_2 & e_1 & 0 \end{bmatrix}$. The 4-vector of Euler parameters is defined as

$\mathbf{p} = [e_0 \quad \mathbf{e}^T]^T$. A direct calculation yields $\mathbf{p}^T \mathbf{p} = e_0^2 + \mathbf{e}^T \mathbf{e} = \cos^2(\chi/2) + \sin^2(\chi/2)\mathbf{u}^T \mathbf{u} = 1$, where the condition that \mathbf{u} is a unit vector has been used. This yields the identity,

$$\mathbf{p}^T \mathbf{p} = 1 \quad (37)$$

which is called the *Euler parameter normalization constraint*.

Numerous identities and derivative relationships involving Euler parameters are due to the quadratic form of terms in the orientation transformation matrix of Eq. (36), as functions of Euler parameters [3]. Defining the 3×4 matrices $\mathbf{E}(\mathbf{p}) := [-\mathbf{e} \quad \tilde{\mathbf{e}} + e_0\mathbf{I}]$ and $\mathbf{G}(\mathbf{p}) := [-\mathbf{e} \quad -\tilde{\mathbf{e}} + e_0\mathbf{I}]$, direct manipulation yields

$$\mathbf{A}(\mathbf{p}) = \mathbf{E}(\mathbf{p})\mathbf{G}(\mathbf{p})^T$$

Carrying out the matrix products indicated shows that

$$\begin{aligned} \mathbf{E}(\mathbf{p})\mathbf{p} &= \mathbf{0} = \mathbf{G}(\mathbf{p})\mathbf{p} \\ \mathbf{G}(\mathbf{p})\mathbf{G}(\mathbf{p})^T &= \mathbf{I} = \mathbf{E}(\mathbf{p})\mathbf{E}(\mathbf{p})^T \end{aligned}$$

where the last relation holds only if \mathbf{p} is normalized. Angular velocity, represented in the body-fixed $x' - y' - z'$ frame, is related to Euler parameters and their derivatives as follows:

$$\begin{aligned} \omega' &= 2\mathbf{G}(\mathbf{p})\dot{\mathbf{p}} \\ \dot{\mathbf{p}} &= \frac{1}{2}\mathbf{G}^T(\mathbf{p})\omega' \end{aligned} \quad (38)$$

Finally, the derivative of $\mathbf{A}(\mathbf{p})\mathbf{a}'$ with respect to \mathbf{p} , with \mathbf{a}' constant, is

$$(\mathbf{A}(\mathbf{p})\mathbf{a}')_{\mathbf{p}} = \mathbf{B}(\mathbf{p}, \mathbf{a}') = 2 \left[(e_0\mathbf{I} + \tilde{\mathbf{e}})\mathbf{a}' \quad \mathbf{e}\mathbf{a}'^T - (e_0\mathbf{I} + \tilde{\mathbf{e}})\tilde{\mathbf{a}}' \right] \quad (39)$$

Author details


Edward J. Haug^{1*} and Petar Pavesic²

1 Department of Mechanical Engineering, The University of Iowa, Iowa City, IA, USA

2 Department of Mathematics, University of Ljubljana and Institute of Mathematics, Physics, and Mechanics, Ljubljana, Slovenia

*Address all correspondence to: echaug@gmail.com

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Petzold LD. Differential/algebraic equations are not ODE's. *SIAM Journal of Scientific and Statistical Computing*. 1982;3(3):367-384
- [2] Warner FE. *Foundations of Differentiable Manifolds and Lie Groups*, Graduate Texts in Mathematics. New York: Springer; 1983
- [3] Haug EJ. *Computer-Aided Kinematics and Dynamics of Mechanical Systems, Modern Methods*. 3rd ed. Vol. II. Berlin: ResearchGate; 2022. Available from: www.researchgate.net
- [4] Strang G. *Linear Algebra and its Applications*. 2nd ed. New York: Academic Press; 1980
- [5] Haug EJ. Extension of Maggi and Kane equations to holonomic dynamic systems. *Journal of Computational and Nonlinear Dynamics*. 2018;13:121003
- [6] Bauchau OA, Laulusa A. Review of contemporary approaches for constraint enforcement in multibody systems. *Journal of Computational and Nonlinear Dynamics*. 2008;3:011005
- [7] Laulusa A, Bauchau OA. Review of classical approaches for constraint enforcement in multibody systems. *Journal of Computational and Nonlinear Dynamics*. 2008;3:011004
- [8] Blajer W. Methods for constraint violation suppression in the numerical simulation of constrained multibody systems-a comparative study. *Computer Methods in Applied Mechanics and Engineering*. 2011;200:1568-1576
- [9] Marques F, Souto AP, Flores P. On the constraints violation in forward dynamics of multibody systems. *Multibody System Dynamics*. 2017;39:385-419
- [10] Stuelpnagel J. On the parametrization of the three-dimensional rotation group. *SIAM Review*. 1964;6(4):422-430
- [11] Pars LA. *A Treatise on Analytical Dynamics*, Reprint by Ox Bow Press (1979). Conn: Woodbridge; 1965
- [12] Atkinson KE. *An Introduction to Numerical Analysis*. 2nd ed. New York: Wiley; 1989
- [13] Tseng F-C, Ma Z-D, Hulbert GM. Efficient numerical solution of constrained multibody dynamics systems. *Computer Methods in Applied Mechanics and Engineering*. 2003;192:439-472
- [14] Garcia de Jalon J, Callejo A, Hidalgo AF. Efficient solution of Maggi's equations. *Journal of Computational and Nonlinear Dynamics*. 2012;7:021003
- [15] Teschl G. *Ordinary Differential Equations and Dynamical Systems*. Providence, Rhode Island: American Math Society; 2012
- [16] Hairer E, Norsett SP, Wanner G. *Solving Ordinary Differential Equations I: Nonstiff Problems*. 2nd ed. Berlin: Springer-Verlag; 1993

Advanced Knotting Techniques (Examples from Surgical Practice)

Kornél Kovách and Daniella Éva Pigniczki

Abstract

In this chapter, the topology of recently used knots is described regarding conventional and laparoscopic surgery. Morphological and technical similarities and differences of the most common knots are discussed, while also considering their transformability. Knots are introduced and made by hand and by instruments, while further demonstrating laparoscopic and microsurgical knotting techniques.

Keywords: surgical knot, knotting technique, knot tying, knot security, conventional knot, laparoscopic knot, microsurgical knot, knot abbreviation, knot notation, composite knot

1. Introduction

Knots are widely used by many professions varying from decorating to life-saving knots and even theoretical ones. Topologists, mathematicians, engineering, sailing and shipping, mountain climbers, and of course surgeons cannot operate without adopting their field's specific knotting techniques. For example, a conventional tied knot (open-loop knot) is not equivalent to the mathematical concept of a knot (closed-loop knot). However, the proper way of tying a functional knot is not so easy to acquire. Therefore, it is essential to clarify the best methodology for this very basic, but also very diversified element of primordial surgical techniques.

A few thoughts as an introduction to the world of knots:

1. A tied knot can easily be converted into a mathematical knot by joining the threads' ends to a closed loop [1]. This transformation can be performed on even an 18-crossing-tied knot. But from the perspective of a mathematical knot, it can hardly be transformed to a conventional tied version from about a 6 or 7 crossing extension (**Figure 1**).
2. There is an enormous difference between mathematical and surgical knots as all tied knots have to count with the laws of Physics such as surface geometry of the thread and consequent torsion, the used material's attributes, friction, elasticity, and producing methodology. Therefore, interprofessional insight is needed to minimize the possible source of mistakes.

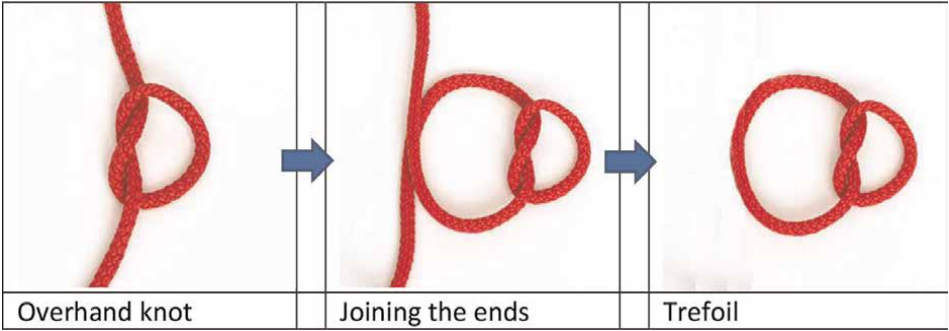


Figure 1.
Convert a tied knot to a mathematical knot.

3. There is a constant common and individual need to improve tying skills in every field of use—especially in the surgical practice and during the so-called learning curve. Thus, limitations and expectations of the different knotting techniques must be well-defined and spread in an accessible form.
4. All professions—especially surgeons—have to set and apply their own standards and nomenclature of their tied knots, which should be clarified during interprofessional communication.
5. Surgeons not only own special tying techniques, but also have specific, sometimes historical names of its knots and knot combinations.
6. We should note that ambient isotopic knots may have a different functioning role in the practice. Mathematically equivalent knots are not necessarily equivalent as conventional tied knots during its practical use.
7. In some cases, only the practitioner’s experience can predict the outcomes and question the theoretic borders of the function of a knot.

2. Detailed explanation

Ad 1: The fact, that composite knots are built up from prime knots, lets us reach a deeper understanding of the structure and the function of knots. This simplified perspective is an essential cornerstone of proper knotting. Furthermore, this simplicity gives space to an easier demonstration of the commonly used 18-crossing knots, that would be more difficult to present with conventional formulas.

Ad 2: Only a small fraction of the theoretically existing knots can be used in practical applications. We can examine a knot’s creating circumstances and conditions, and investigate its final functions and limitations as well.

Knots can be made for countless reasons from poorly functional decorative knots through the climbing knots to surgical knots that have serious practical importance. The most common function of a knot besides decoration is to fix objects and get them stick together. Fixable things can have different characteristics such as shape, flexibility, rigidity, and capability of resistance. Fixation is usually provided by the loop of

a cord or a thread secured by knots. The knot-tying procedure can be performed under tension or not under tension, where the previously fastened knot suffers the burden later. Selecting the proper knotting technique depends on the circumstances of the tying process and the properties of the participating objects. Untyability is also a crucial point of view during the selection of the appropriate type of the knot—like the untying property is surely needed in case of sailing and is unacceptable during surgical procedures.

The most frequent knot application (referring to surgery):

- a. Strangulation of a tubular object or organ (e.g., closure of a sac or an open vessel): the first knot of the loop strangulates while the following ones strengthen the participant while the whole tying process is performed under tension.
- b. Skin stitches, for example, vertical mattress stitch: adjustment of the wound edges in the same plane tight and under tension.
- c. Bowline knot: previously blinded knot (not under tension) for postponed application and strain. In surgical practice, we decisively use under-tension knotting techniques for holding and fixing the living elastic tissues together. Thus, the loop-fixing knots must be safe and hard or impossible to untie. Our proper tying techniques are discussed in a later section.

Ad 3: Numerous ways exist to describe mathematical knots. The most common methods are planar diagrams, Gauss codes, Morse link presentation, braid representation, etc. However, it can be trying to find parallelism between these methods regarding the similarly made knots or the more complex composite knots. It is also difficult to prepare a tied knot based on mathematical nomenclature. Surgical knots are mostly made of numerous simple prime knots following each other in a well-defined order. Therefore, it was easy to create a specific nomenclature which is representative and provides a proper description of the composite knots' appearance. This gives us a simple way to compare the bearing capacity and the nature of different knots.

3. Special notation and abbreviations in surgical knotting

3.1 The base “half knot”

The root of surgical knotting is the single twisted half knot. Mathematically, the half knot is a disjunct trefoil (AB 3.1) knot. The notation of this half knot is “1,” or H1. (Other general notation for that term is “overhand knot”) (**Figure 2**).

The half knot can also be formed by more than one throws. If it is formed by two throws, the notation is “2,” or H2. This knot is a disjunct cinquefoil knot (AB 5.1) (or “double overhand knot”) (**Figure 3**).

Three, or more throws in a half knot is uncommon (“3” or H3, and above). These simple or “opened” prime knots are the basics of the practical used composite knots.

In these knots, both threads enter to the crossing point from the same direction, bending and twisting through an imaginary line, and leave it to a continuous direction.

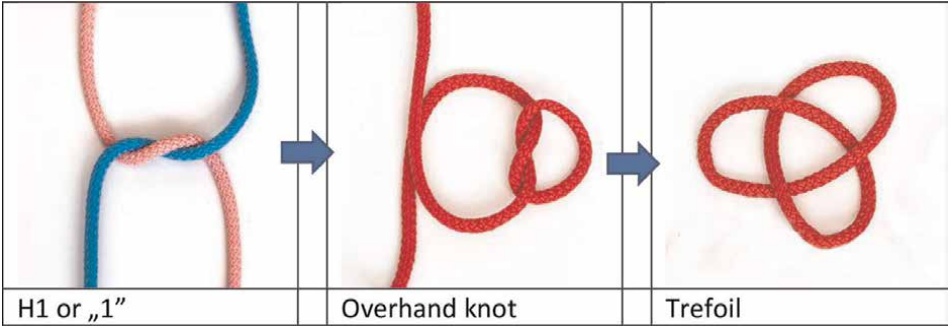


Figure 2.
H1 knot transformation.

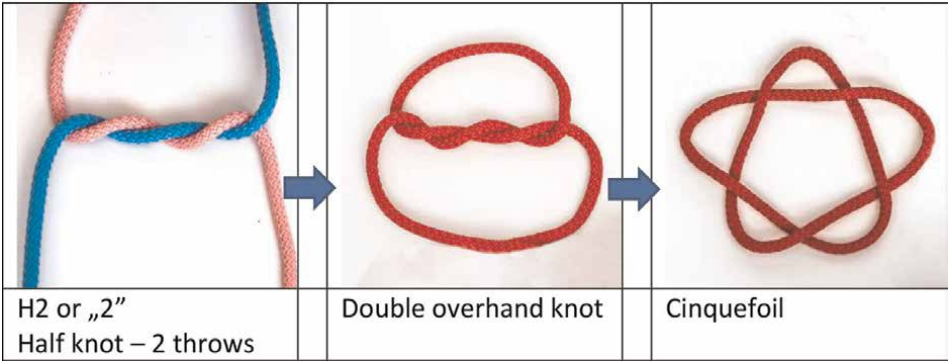


Figure 3.
H2 knot transformation.

3.2 The base half hitch

In the surgical notation, there are equivalent pairs of these knots, which are invariants (ambient isotopy), but have different initial shapes, and characteristics.

This ambient isotopic pair of the half knot is the half hitch, or otherwise called the sliding knot. The notation of this knot is “S.” In this case, one part of the thread is active and the other is passive. The passive thread is nearly straight while the active thread ends it is turn around the passive one. Therefore, the active end leaves the knot toward the opposite direction than the entering (**Figure 4**).

If the active end turns around more than one time, the notation is “2S,” or S2, or above (**Figure 5**).

3.3 Linking knots

From the combination of these basic knots, sequential practical knots (composite knots) are derived.

3.4 Combining half knots

There are two possibilities to combine half knots (“1, H1”). Using a symmetric knot means that the threads are entering to the linking loop from the same side (both

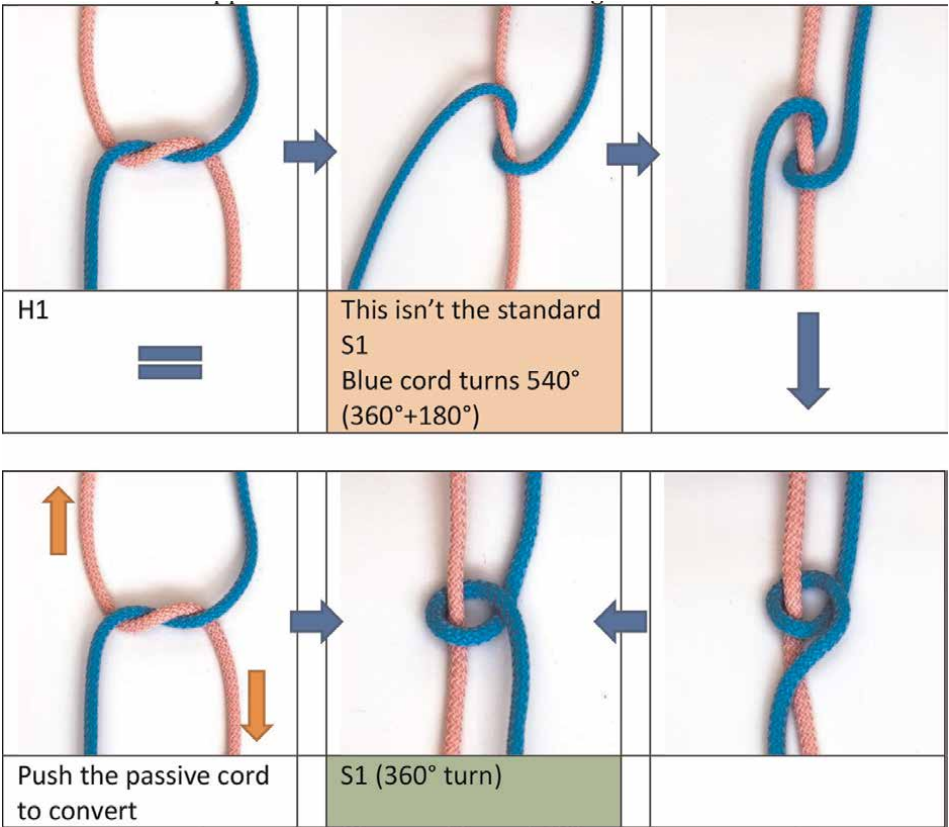


Figure 4.
S1 knot, H1 to S1 transformation.

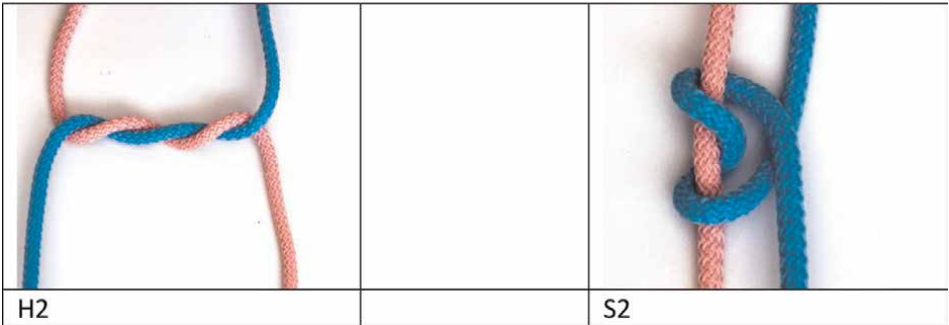


Figure 5.
S2 knot, H2S2 transformation.

under or above), while in the case of the asymmetric knot, the ends are entering from the opposite direction (one from above and one from under). The notation of symmetrically linking half knots is “=” or “s” (e.g., 1 = 1 or H1H1s). The notation of asymmetrical knots is “×” or “a” (e.g., 1 × 1 or H1H1a). The 1 = 1 (H1H1s) is called the square knot (reef knot, surgical knot {not surgeon’s knot}), which is a knot consisting of two trefoils with opposite chirality’s. The 1 × 1 (H1H1a) is the granny’s knot, which is the knot consisting of two trefoils with the same chirality (Figure 6).

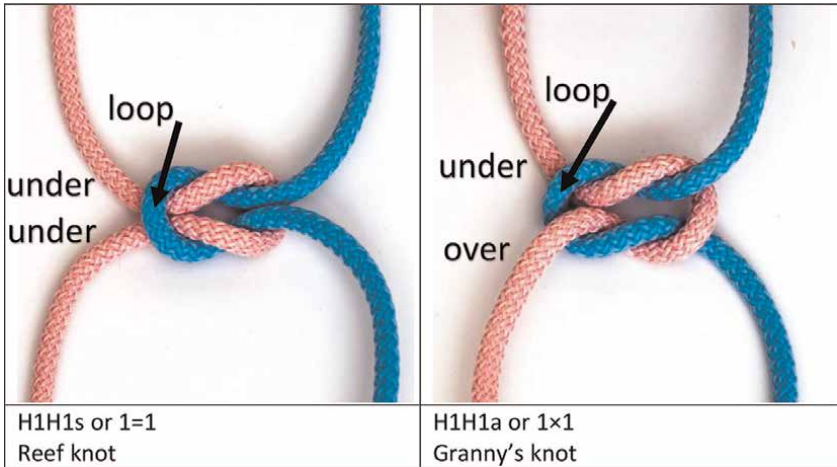


Figure 6.
Flat knots.

Based on that, it is easy to present the classic surgeon's knot which is described by the formula of $2 = 1$ (H2H1s). Generally used safe knots in surgery is $1 = 1 = 1 = 1 = 1$ (H1H1sH1sH1sH1s), or $2 = 2 (= 1 = 1)$ (H2H2s{H1sH1s}) (**Figure 7**).

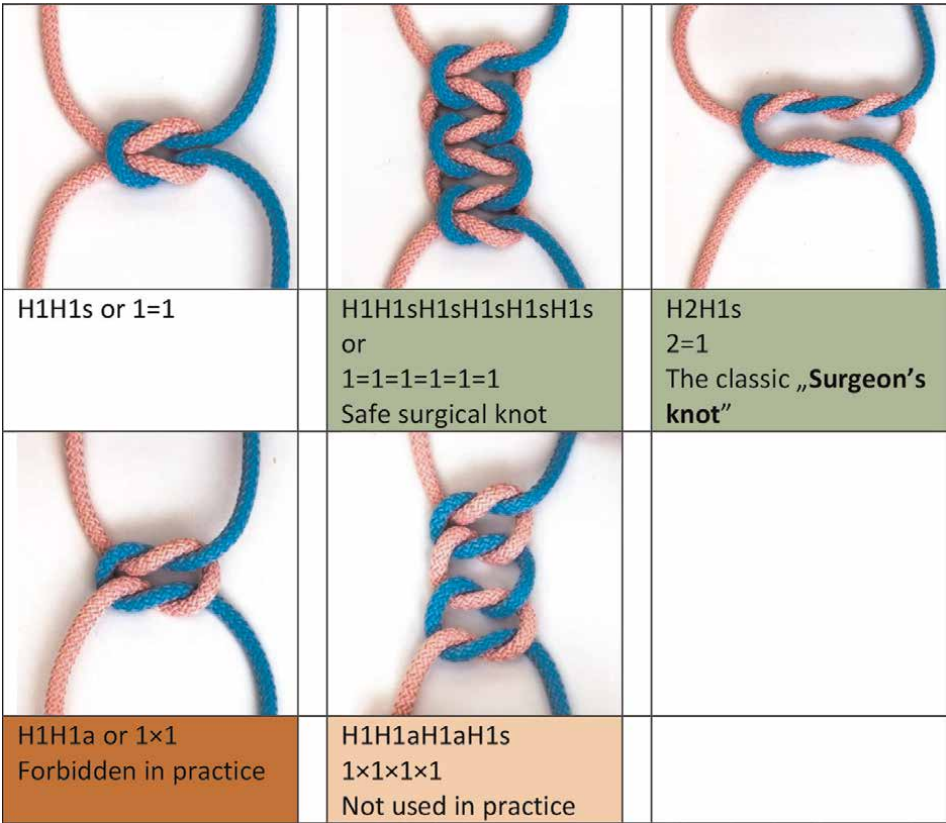


Figure 7.
Some types of flat knots.

3.5 Combining half hitches: sliding

To form and combine half hitches, the first key step is to identify the relation between the passive and the entering active end. The passive thread is the one which lays under or above the entering active end (depending on the type). If the direction of the entering thread at the crosspoint changes from half hitch to half hitch, the knot sequence is symmetric. If it keeps crossing from the same direction, the knot combination is asymmetric. The notation of the symmetric combination is “ \times ” or “s” (e.g., $S \times S$ or SSs). The notation of the asymmetric sequence is “=” or “a” (e.g., $S=S$, SSa or can be left: SS(a) or SS) (Figures 8 and 9).

3.6 Combining half hitches—blocking

When half hitches are combined, there is another important possibility to define function. If the passive thread remains untouched after the tying process, it could suffer continuous sliding. But fortunately, it can be blocked if we perform a change on the passive thread. The notation of the half hitches’ blocking sequence is “//” or “b” (e.g., $S//S$ or SSb). Therefore if “//” or “b” is missing, the knot is a continuous sliding knot. We also have to mark the symmetry in the case of blocking knots. In case of symmetry we use the plan blocking sign “//” or “b” (clarified sign is: SSsb, e.g., $S//S$, SSb or SSsb). In case of an asymmetric blocking combination, the notation is “// \times ”, “ab” or “(a)b” (e.g., $S//\times S$, SS(a)b or SSab) (Figures 10 and 11).

Ad 4. Knot conversion. Most important thing is to be aware of the properties of the used knots for the exact task. We also have to know which knots are equivalent and how to convert them if necessary [2]. For example, in a place which is hard to access, we should prefer sliding knots and then transform it to a blocking knot. The next table presents the equivalency of the different knots.

The $1 = 1$ (H1H1s) knot can be transformed to an $S \times S$ (SSs) or an $S//S$ (SSb) knot.
The 1×1 (H1H1a) knot can be transformed to an $S=S$ (SS(a)) or an $S//\times S$ (SSab) knot (Figure 12).

This notation leaves the symmetricity index unchanged throughout the conversions ($H1H1s > SSs > SSsb$, or $H1H1a > SSa > SSab$). It can be distractable, that the more frequently used abstract notation is changing during knot conversions ($1 = 1 > S \times S > S//S$), ill. ($1 \times 1 > S=S > S//\times S$) and it is not a mistake but a habitual use.

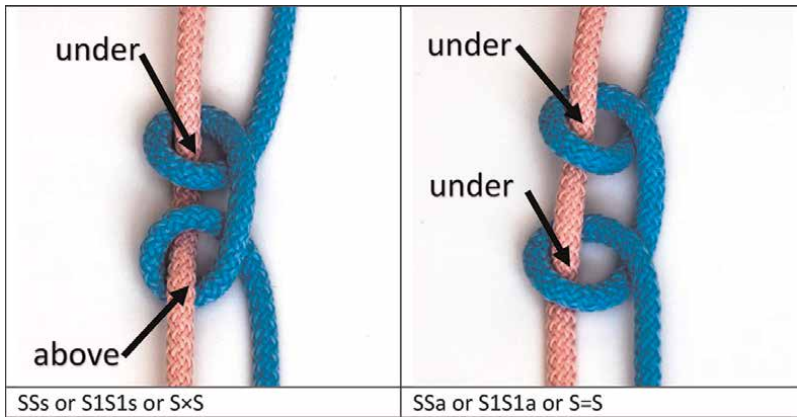


Figure 8.
Basic sliding knots.

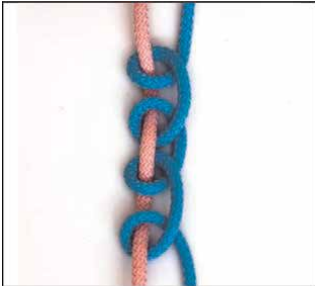


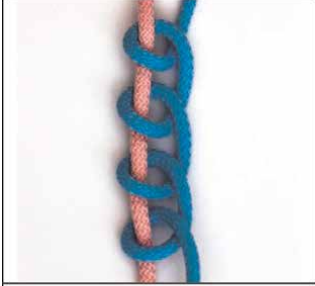


		
SSsSsSs or S×S×S×S or S1SsS1sS1sS1s	S2S1s or 2S×S	S2S2s or 2S×2S
		
SSaSaSa or S=S=S=S or SSSS or S1aS1aS1aS1a	S2S1a or 2S=S	S2S2a or 2S=2S

Figure 9.
Some examples of sliding knots.

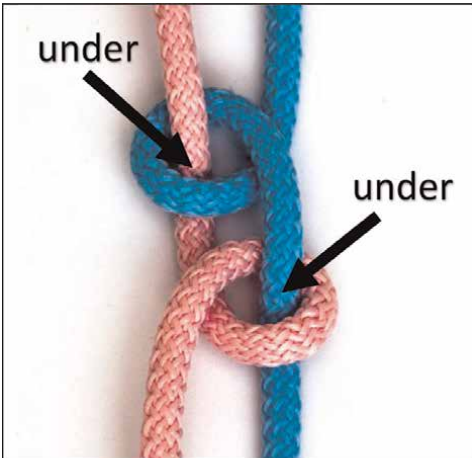
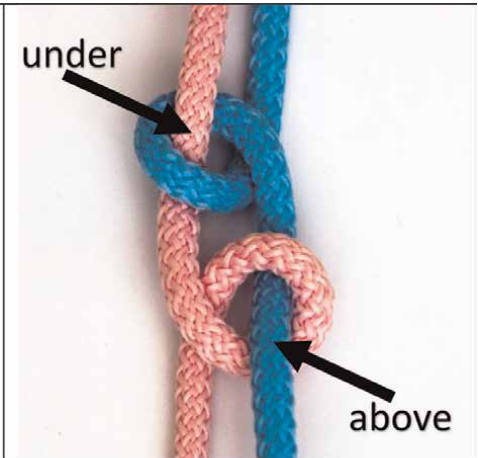
	
SSb or SSsb or S//S or S1S1sb	SSab or S//×S or S1S1ab

Figure 10.
Basic blocking knots.

Ad 5. knotting, dynamic knotting technics. To use the most suitable knot for a task, not only the awareness of the knot's conformation, but also the tying method is crucial. A knot can have two basic functional use during practical application. On one hand, we can bind knot for postponed application and strain it later (e.g., some mountain climbing knots). On the other hand, the knot can be tied under tension,

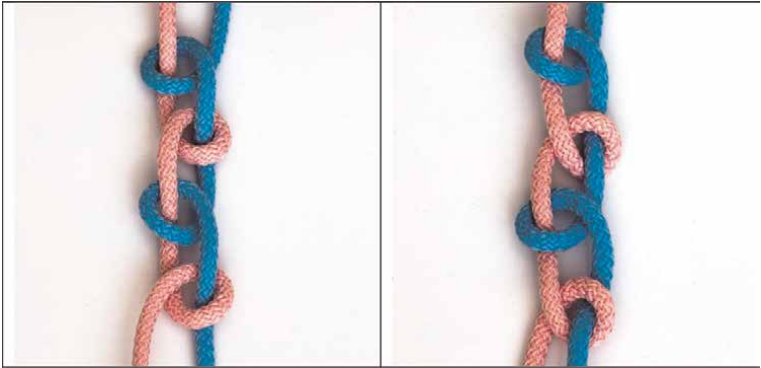


Figure 11.
More blocking knots.

H1H1s or 1=1	S1S1s or SxS	S1S1sb or S//S
H1H1a or 1x1	S1S1a or S=S	S1S1ab or S//xS
H2H1s or 2=1	S2S1s or 2SxS	
Plan knot	Sliding knot	Blocking knot

Figure 12.
Knot conversation.

when even the initial form of the knot suffers strain that will expand by the end of the process. This ulterior case demands a more complex approach. The attachable living tissue elements are usually elastic, slippery and even moving. All these properties facilitate the untie of the knot and the consequent release of the tissue. Only the knot itself can maintain the attaching effect, and therefore the knot sequence should be built up with a constant holding technique.

Tense-holding methods are:

- a. the first half knot should be tied tense and then the following knots should be made with the constant pulling of the thread ends.
- b. we can perform a sliding knot that will be either transformed into a blocking knot or stressed among the active and the passive thread.
- c. while making the first half knot, we can perform more than one throws and therefore there will be a bigger frictional resistance between the threads that can hold the half knot together tightly.
- d. the first half knot can be grabbed by assisting tool until we place the next half knot on the top of it.

Ad a) a knot can be made by hands or with instruments (apodactylic knotting technique) [3, 4]. Suggested video: <https://www.youtube.com/watch?v=SL3lF17ocuE>. Hand-knotting techniques have their own science and notation (one-handed, two-handed tying technique, instrument-tied technique, etc.). We have to note, that only the two-handed knotting technique is suitable for knotting under tension from all possibilities (**Figures 13–15**).

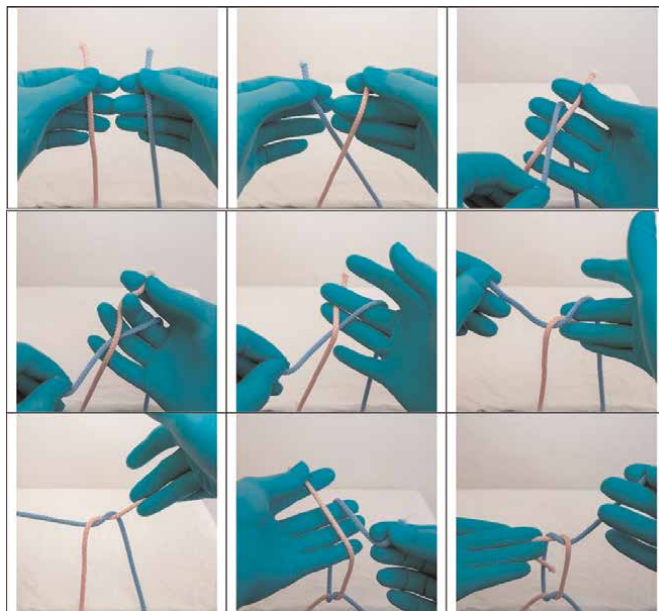


Figure 13.
One-hand tie technique—alternating hands.

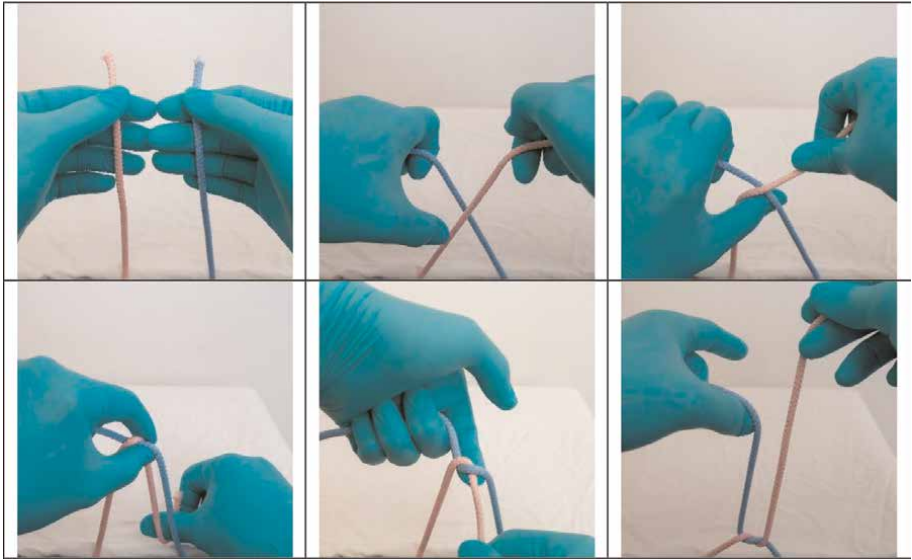


Figure 14.
Two-hand tie technique—first H1 half knot.

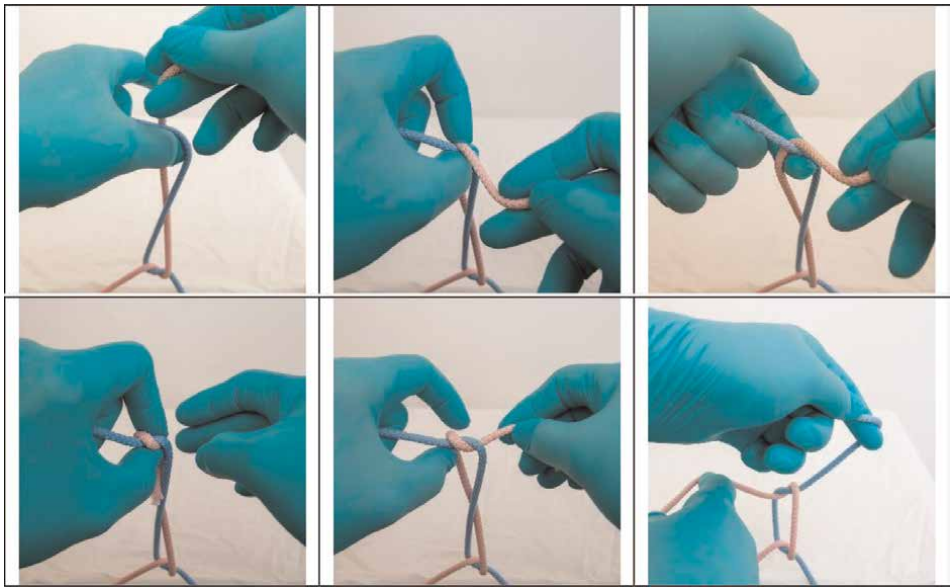


Figure 15.
Two-hand technique—second symmetric H1 knot pair.

Ad b) (continuous) sliding knots were previously discussed that can assure continuous tension. Here we present some sliding knots usually employed in Orthopedics [5]. These knots receive a really great burden, therefore these should hold especially tight (**Figure 16**).

Ad c) multiple throws applied in case of the first half knot are setting a larger joining surfaces. The modeling of this is a very complex task and depends on elasticity, the thread's composition (mono- or multifilament) and frictional resistance. These attributes of the knots are widely investigated by numerous research groups [6]. In

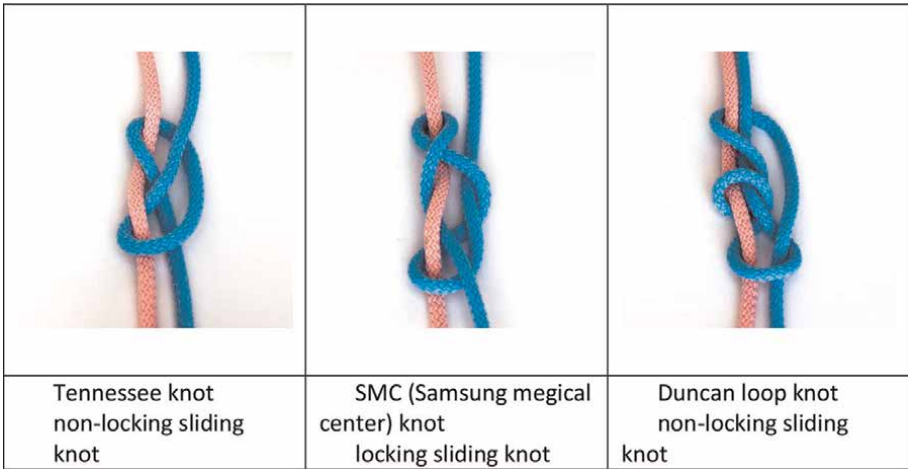


Figure 16.
Sliding (arthroscopic) knots.

surgical practice, mostly the double-throwed first half knot is used and enough, more throws are barely used.

Ad d) the first half knot can be grabbed and fixed by surgical forceps and released right just before the tightening of the second half knot.

4. Knot security

The tying technics are to ensure the security of the knots. When talking about the security of a surgical knot we distinguish loop security and knot security.

The loop security means that during the knotting the distance between the tied tissues at the first loop is not increasing, therefore the diameter of the loop is not growing and therefore the loop will not get loose. The knot security means that the knot ensures a safe closure, thus the knot sequence does not or only due to a really strong force can a) untie, b) slide, c) break. In research regarding knot testing usually the knot security defined and investigated as it is more standardized. By contrast, the loop security only provides some additional information and is investigated as an acceptance interval measured in mm.

The secure knots (**Table 1**) [7].

	Secure							
+	H1s		Sb					Sb
+	H1s		S					Sb
+	H1s		H1s					S Sb
+	H1s	H1a	H1s	H1a	H2s	H2a	H2s	H2a S
	H1		H2		H2		H3	
Basis	Unsafe		Unsafe				Safe	Safe

Table 1.
Composite knot securing.

A challenge is rising when we need to tie a knot where there is not enough space for our hands (or the hand-knotting is too rough). For example, knotting in deep cavity or knotting with minimally invasive techniques (laparoscopy, arthroscopy, thoracoscopy, and microsurgery). In such cases, instrumental knotting techniques are used. This method has two fundamental subgroups: extracorporal- and intracorporal-knotting techniques. During extracorporal knotting, we make the knot at the free space (extracorporally) and then insert it into the badly accessible final spot with an instrument. Usually during extracorporal knotting we tie a sequenced sliding knot, then add additional fixing knots, prepared as blocking knots that are also pushed down as SSb knots. These knots can be made even through a 5 mm wide laparoscopic working tunnel (port). Two tools are used during intracorporal knotting.

4.1 Instrumental knotting techniques

See **Figure 17.**

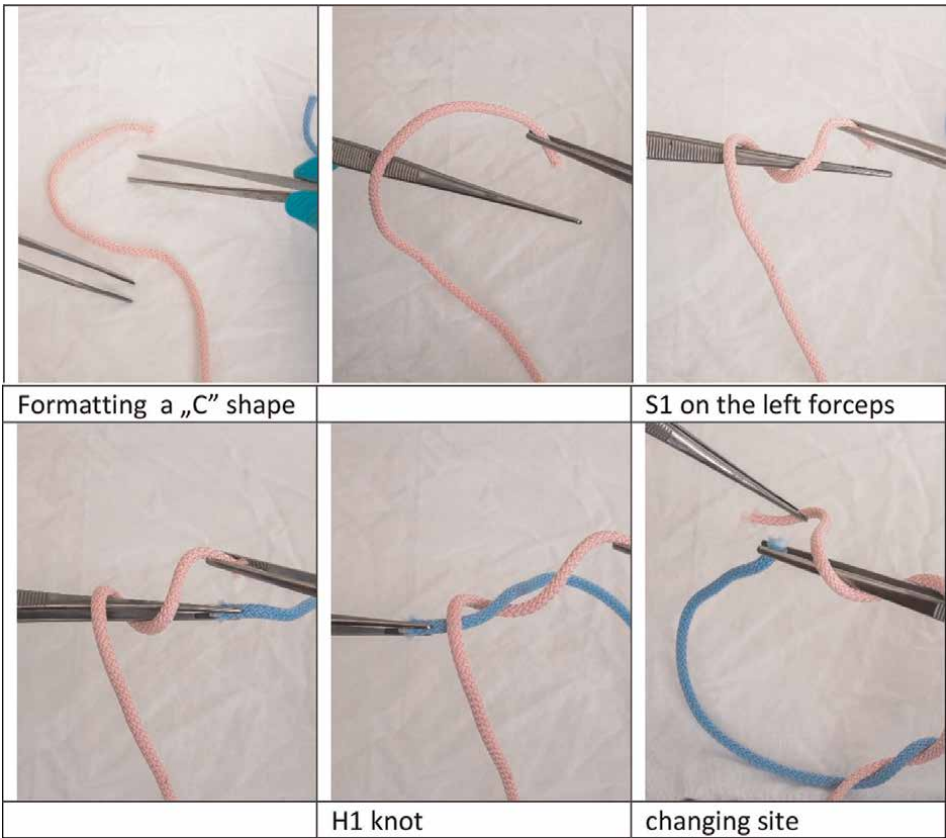


Figure 17.
Instrumental knotting.

5. Conclusions


There are several nomenclatures for the composite knots made of the combination of simple prime knots. It is suggested to use the detailed nomenclature which includes not only the type of the knots (H, S) but also the number of throws (e.g., H1, H2, H3, S1, S2, etc.). In the case of composite knots, it is suggested to use the symmetry/asymmetry markings which can constantly be used even in case of conversions. Instead of the “=“ ↔ “×”, identical ↔ non-identical nomenclature, the “s” (symmetrical), and the “a” (asymmetrical) index are suggested. There is great potential in the partnership of topology as it can provide a deeper understanding through the investigation of the dynamic behavior of knots—and that field has to be further studied.

Author details

Kornél Kovách* and Daniella Éva Pigniczki
Department of Surgery, University of Szeged, Szeged, Hungary

*Address all correspondence to: kovach.kornel@med.u-szeged.hu

IntechOpen

© 2023 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] Adams CC. *The Knot Book: An Elementary Introduction to the Mathematical Theory of Knots*. New York, NY: W. H. Freeman and Company; 2001
- [2] Zhang W, Wu X. How to convert a square knot or surgeon's knot into a sliding knot and an assessment of their sliding and re-locking properties? *The Surgeon*. 2020;**18**(6):327-334. ISSN 1479-666X. DOI: 10.1016/j.surge.2019.11.004
- [3] Giddings FD. *Surgical Knots and Suturing Techniques*. 5th ed. Anacortes WA: Giddings Studio; 2018. 12-23 p. ISBN: 9781984007537
- [4] Kirk RM. *Basic Surgical Techniques*. 6th ed. Churchill Livingstone; 2010. pp. 1-217. ISBN: 9780702033919
- [5] Akindele RA, Fasanu AO, Mondal SC, Komolafe JO, Mishra RK. Comparing extracorporeal knots in laparoscopy using knot and loop securities. *World Journal of Laparoscopic Surgery*. 2014; **7**(1):28-32
- [6] Jennifer Chu: Untangling the mechanics of knots: New model predicts the force required to tie simple knots. Available from: <https://news.mit.edu/2015/untangling-mechanics-knots-0908>. [Accessed: 2022-12-30]
- [7] Romeo A, Rocha CL. What is the best surgeon's knot? Evaluation of the security of the different laparoscopic knot combinations. *Journal of Minimally Invasive Gynecology*. 2018;**25**(5): 902-911. DOI: 10.1016/j.jmig.2018.01.032 Epub 2018 Feb 5



Edited by Paul Bracken

Topology remains an active and fundamental area of research that plays a foundational role in many branches of mathematics and science, such as analysis, differential geometry, physics and even biology. It is hoped the papers in this book will contribute to stimulating research in this basic area of mathematics.

Published in London, UK

© 2023 IntechOpen

© Paul Campbell / iStock

IntechOpen

