A thesis concerning symbolic sequence processing

version 0.50e

Daniel J. Greenhoe
(inside front cover)
A thesis concerning
symbolic sequence processing

Daniel J. Greenhoe (柯晨光)
This text was typeset using X\TeX, which is part of the \TeX family of typesetting engines, which is arguably the greatest development since the Gutenberg Press. Graphics were rendered using the \texttt{pstricks} and related packages, and \LaTeX graphics support. The main roman, \textit{italic}, and \textbf{bold} font typefaces used are all from the \textit{Heuristica} family of typefaces (based on the \textit{Utopia} typeface, released by \textit{Adobe Systems Incorporated}). The math font is XITS from the XITS font project. The font used in quotation boxes is adapted from \textit{Zapf Chancery Medium Italic}, originally from URW++ Design and Development Incorporated. The font used for the text in the title is \textit{Adventor} (similar to \textit{Avant-Garde}) from the \TeX-Gyre Project. The font used for the version in the footer of individual pages is \texttt{Liquid Crystal} (\textit{Liquid Crystal}) from FontLab Studio. The Latin handwriting font is \textit{Lavi} from the \textit{Free Software Foundation}. Chinese glyphs appearing in the text are from the font 王漢宗中明體繁.\textsuperscript{1}

The ship appearing throughout this text is loosely based on the \textit{Golden Hind}, a sixteenth century English galleon famous for circumnavigating the globe.

\textsuperscript{1}pinyin: Wáng Hán Zōng Zhōng Míng Tí Fán; translation: Hán Zōng Wáng's Medium-weight Míng-style Traditional Characters; literal: 王漢宗~font designer's name; 中~medium; 明~Míng (a dynasty); 體~style; 繁~traditional
“Here, on the level sand,
Between the sea and land,
What shall I build or write
Against the fall of night?”

“Tell me of runes to grave
That hold the bursting wave,
Or bastions to design
For longer date than mine.”

Alfred Edward Housman, English poet (1859–1936) ²

“The uninitiated imagine that one must await inspiration
in order to create. That is a mistake. I am far from saying
that there is no such thing as inspiration; quite the oppo-
site. It is found as a driving force in every kind of human
activity, and is in no wise peculiar to artists. But that
force is brought into action by an effort, and that effort
is work. Just as appetite comes by eating so work brings
inspiration, if inspiration is not discernible at the begin-
ning.”

Igor Fyodorovich Stravinsky (1882–1971), Russian-born composer ³

“As I think about acts of integrity and grace, I realise that
there is nothing in my knowledge to compare with Frege's
dedication to truth. His entire life's work was on the verge
of completion, much of his work had been ignored to the
benefit of men infinitely less capable, his second volume
was about to be published, and upon finding that his fund-
amental assumption was in error, he responded with intel-
lectual pleasure clearly submerging any feelings of personal
disappointment. It was almost superhuman and a telling
indication of that of which men are capable if their dedi-
cation is to creative work and knowledge instead of cruder
efforts to dominate and be known.”

Bertrand Russell (1872–1970), British mathematician, in a 1962 November 23 letter to Dr. van Heijenoort. ⁴
ACKNOWLEDGEMENTS

It is not far from the truth to say that much of the research presented in this text started with this email from my academic advisor, Professor Po-Ning Chen:\(^5\)

```
Sat, 25 Jan 2014 03:45:51 -0800 (PST)
Dear Dan:  So far people are mostly (and are used to) dealing with “frequency-transform” of a numerical sequences. However, sometimes, we need to find out the “occurrence frequency” of a symbolic sequence. I am wondering “Can we design a wavelet transform for a symbolic sequence like DNA?” What do you think? See the attached paper as an example.
Po-Ning
```

I responded the same day saying, “Thank you for the paper. So far I have read the introduction. I have never heard of “Genomic Signal Processing” before. That is very interesting, and maybe something that could help a lot of people some day....I will investigate Genomic Signal Processing more. Thank you for introducing me to the topic.”\(^6\)

This text is not about genomic signal processing in particular. But it is about symbolic sequence processing, of which genomic signal processing is a special and motivational case.

So I would like to say one more time to my academic advisor, Professor Po-Ning Chen, “Thank you for introducing me to the topic”(!) which has to a large extent resulted in the research described in this text.

Moreover, it would be difficult to effectively communicate the ideas in this text without the use of the typesetting software XeLaTeX, and \(\LaTeX\) upon which it is based. Even though I am from the United States and \(\LaTeX\) has its origins in the United States, it seems I never really knew what it was before coming to Taiwan and being introduced to it by Professor Chen. And likewise XeLaTeX has it’s origins in the United States as well, but again I was introduced to it in Taiwan, this time by Professor Peng-Hua Wang.\(^7\) So I would like to thank Professor Chen for introducing me to \(\LaTeX\), and I would like to thank Professor Wang for introducing me to XeLaTeX.

Finally, I would like to thank National Chiao-Tung University (NCTU), the Communications Engi-

---

\(^5\)Po-Ning Chen (Chinese: 陈伯宁, pinyin: Chén Bó Níng)

\(^6\)The “attached paper” referred to by Professor Chen’s email is \(\text{Galleani and Garelo (2010).}\)

neering Department of NCTU, and the people of Taiwan. Although born and raised in the United States, Taiwan has graciously allowed me to live, study, and work here for several years now. To live in a country not one’s own is always a privilege, an honor, and a great opportunity to learn about life. I am very happy to have had and to continue to have this privilege.

—Daniel J. Greenhoe (柯晨光)
Contents

Front matter
   Title page .................................................. i
   Typesetting information ................................... ii
   Quotes ....................................................... iii
   Acknowledgements .......................................... v
   Contents ................................................... vii

1 Foundations ............................................... 1
   1.1 Introduction ............................................. 1
   1.2 Further mathematical support ............................ 3
   1.3 Standard definitions ..................................... 4
      1.3.1 Sets .................................................. 4
      1.3.2 Relations ............................................. 5
      1.3.3 Field operator pairs ................................. 7
      1.3.4 Graph Theory ......................................... 7
   1.4 Traditional probability ................................... 8
   1.5 Order space concepts .................................... 10
      1.5.1 Order ............................................... 10
      1.5.2 Lattices .............................................. 11
      1.5.3 Isomorphic spaces .................................... 14
      1.5.4 Monotone functions on ordered sets ................. 16
   1.6 Metric space concepts ................................... 20
      1.6.1 Motivation ............................................ 20
      1.6.2 Isometric spaces ..................................... 20
   1.7 Ordered metric spaces ................................... 22
      1.7.1 Definitions .......................................... 22
      1.7.2 Examples ............................................. 23
   1.8 Sequences ................................................ 23
      1.8.1 Sequences ............................................ 23
      1.8.2 Filtering ............................................. 25
      1.8.3 Discrete Fourier Transform .......................... 26

2 Stochastic processing on weighted graphs ................ 27
   2.1 Outcome subspaces ....................................... 27
      2.1.1 Definitions .......................................... 27
      2.1.2 Specific outcome subspaces .......................... 28
      2.1.3 Example calculations ................................ 29
   2.2 Random variables on outcome subspaces ................. 46
      2.2.1 Definitions .......................................... 46
      2.2.2 Properties ........................................... 47
      2.2.3 Problem statement .................................... 49
      2.2.4 Examples ............................................. 49
   2.3 Operations on outcome subspaces ....................... 66
      2.3.1 Summation ............................................ 66
      2.3.2 Multiplication ....................................... 73

vii


<table>
<thead>
<tr>
<th>2.3.3 Metric transformation</th>
<th>..........................................................</th>
<th>74</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Symbolic sequence processing on RN</td>
<td>77</td>
<td></td>
</tr>
<tr>
<td>3.1 Outcome subspace sequences</td>
<td>..........................................................</td>
<td>77</td>
</tr>
<tr>
<td>3.1.1 Definitions</td>
<td>..........................................................</td>
<td>77</td>
</tr>
<tr>
<td>3.1.2 Examples of symbolic sequence statistics</td>
<td>..........................................................</td>
<td>78</td>
</tr>
<tr>
<td>3.2 Extending to distance linear spaces</td>
<td>..........................................................</td>
<td>82</td>
</tr>
<tr>
<td>3.2.1 Motivation</td>
<td>..........................................................</td>
<td>82</td>
</tr>
<tr>
<td>3.2.2 Some random variables</td>
<td>..........................................................</td>
<td>82</td>
</tr>
<tr>
<td>3.2.3 Some ordered distance linear spaces</td>
<td>..........................................................</td>
<td>83</td>
</tr>
<tr>
<td>3.3 Symbolic sequence processing applications</td>
<td>..........................................................</td>
<td>85</td>
</tr>
<tr>
<td>3.3.1 Low pass filtering/Smoothing</td>
<td>..........................................................</td>
<td>85</td>
</tr>
<tr>
<td>3.3.2 High pass filtering</td>
<td>..........................................................</td>
<td>91</td>
</tr>
<tr>
<td>3.3.3 Fourier Analysis</td>
<td>..........................................................</td>
<td>95</td>
</tr>
<tr>
<td>3.3.4 Wavelet Analysis</td>
<td>..........................................................</td>
<td>100</td>
</tr>
</tbody>
</table>

A Lagrange arc distance | 105 |
| A.1 Introduction | .......................................................... | 105 |
| A.1.1 The spherical metric | .......................................................... | 105 |
| A.1.2 Linear interpolation | .......................................................... | 106 |
| A.1.3 Polar linear interpolation | .......................................................... | 106 |
| A.1.4 Distance in terms of polar linear interpolation arcs | .......................................................... | 108 |
| A.2 Definition | .......................................................... | 108 |
| A.3 Calculation | .......................................................... | 109 |
| A.4 Properties | .......................................................... | 112 |
| A.4.1 Arc function R(p,q) properties | .......................................................... | 112 |
| A.4.2 Distance function d(p,q) properties | .......................................................... | 116 |
| A.5 Examples | .......................................................... | 120 |

B Distance spaces | 123 |
| B.1 Introduction and summary | .......................................................... | 123 |
| B.2 Fundamental structure of distance spaces | .......................................................... | 125 |
| B.2.1 Definitions | .......................................................... | 125 |
| B.2.2 Properties | .......................................................... | 125 |
| B.3 Open sets in distance spaces | .......................................................... | 126 |
| B.3.1 Definitions | .......................................................... | 126 |
| B.3.2 Properties | .......................................................... | 126 |
| B.4 Sequences in distance spaces | .......................................................... | 128 |
| B.4.1 Definitions | .......................................................... | 128 |
| B.4.2 Properties | .......................................................... | 129 |
| B.5 Examples | .......................................................... | 131 |

C Power distance spaces | 135 |
| C.1 Definitions | .......................................................... | 135 |
| C.2 Properties | .......................................................... | 136 |
| C.2.1 Relationships of the power triangle function | .......................................................... | 136 |
| C.2.2 Properties of power distance spaces | .......................................................... | 137 |
| C.3 Examples | .......................................................... | 143 |

D Tools from analysis | 147 |
| D.1 Linear spaces | .......................................................... | 147 |
| D.1.1 Structure | .......................................................... | 147 |
| D.1.2 Metric Linear Spaces | .......................................................... | 148 |
| D.1.3 Normed Linear Spaces | .......................................................... | 149 |
| D.1.4 Relationship between metrics and norms | .......................................................... | 149 |
| D.2 Metric spaces | .......................................................... | 153 |
| D.2.1 Algebraic structure | .......................................................... | 153 |
| D.2.2 Metric preserving functions | .......................................................... | 156 |
| D.2.3 Product metrics | .......................................................... | 159 |
| D.3 Sums | .......................................................... | 161 |
| D.3.1 Summation | .......................................................... | 162 |
| D.3.2 Convexity | .......................................................... | 163 |
D.3.3 Power means .......................................................... 163
D.3.4 Inequalities .......................................................... 167
D.4 Topological Spaces .................................................... 168

E Interpolation ............................................................ 173
E.1 Lagrange interpolation .............................................. 173
E.2 Newton interpolation ............................................... 174

F C++ source code support ............................................. 177
F.1 Symbolic sequence routines ....................................... 177
F.2 Die routines .......................................................... 182
F.3 Real die routines ..................................................... 187
F.4 Spinner routines ...................................................... 193
F.5 DNA routines ........................................................ 199
F.6 Le Grange arc distance routines ............................... 208

Back Matter .............................................................. 215
References ............................................................... 215
Reference Index ........................................................ 237
Subject Index ............................................................ 241
End of document ......................................................... 256
CHAPTER 1

"Je me plaisois surtout aux mathématiques, à cause de la certitude et de l'évidence de leurs raisons: mais je ne remarquois point encore leur vrai usage; et, pensant qu'elles ne servoient qu'aux arts mécaniques, je m'étonnois de ce que leurs fondements étant si fermes et si solides, on n'avait rien bâti dessus de plus relevé."

"I was especially delighted with the mathematics, on account of the certitude and evidence of their reasonings; but I had not as yet a precise knowledge of their true use; and thinking that they but contributed to the advancement of the mechanical arts, I was astonished that foundations, so strong and solid, should have had no loftier superstructure reared on them."

René Descartes, philosopher and mathematician (1596–1650)

1.1 Introduction

In traditional stochastic processing, a traditional random variable $X$ first maps the underlying stochastic process to the real line, and then operations (such as the expected value operation $E(X)$) is performed on $X$. Here, real line refers to the structure $(\mathbb{R}, |\cdot|, \leq)$, where $\mathbb{R}$ is the set of real numbers, $\leq$ is the standard linear order relation on $\mathbb{R}$, and $d(x, y) \triangleq |x - y|$ is the usual metric on $\mathbb{R}$.

This is all well and good when the physical process being analyzed (in the case of statistical estimation or system analysis) or being processed (as in the case of signal processing, including digital signal processing) is also linearly ordered and has a metric geometry similar to the one induced by the usual metric on $\mathbb{R}$.

Be that as it may, in several real world applications, this is simply not the case. Take these processes for example:

---

1 quote: Descartes (1637a)
translation: Descartes (1637b) (part I, paragraph 10)
The values of a *fair die* \{□, △, ○, ◆, ◆, ◆\} have absolutely no order structure, and have no metric except the *discrete metric*. On a fair die, □ is not greater or less than △; rather □ and △ are simply symbols without order. Moreover, □ is not “closer” to △ than it is to ○; rather, □, ○, and △ are simply symbols without any inherit order or metric geometry.

*Genomic Signal Processing* (GSP) analyzes biological sequences called *genomes*. These sequences are constructed over a set of 4 symbols that are commonly referred to as □, △, ○, and ◆, each of which corresponds to a nucleobase (adenine, thymine, cytosine, and guanine, respectively). A typical genome contains a large number of symbols (about 3 billion for humans, 29751 for the SARS virus).

A *linear congruential pseudo-random number generator* induced by the equation \(y_{n+1} = (y_n + 2) \mod 5\) with \(y_0 = 1\). The sequential nature of the structure induces both a natural order and distance.

In all three processes, the symbols have an order structure and a *metric geometry* that is fundamentally dissimilar from that of the *real line*. Therefore, statistical inferences based on the *real line* will likely lead to results that arguably have little relationship with intuition or reality.

So we can observe the following:

1. A traditional random variable \(X\) maps to the real line.
2. The structure of the underlying stochastic process (the domain of \(X\)) may be very dissimilar to that of the real line.
3. To fix the problem, we need random variables that map to alternative structures that are more similar to the underlying stochastic processes.
4. Such a structure should have both an *order relation* and a *distance function* defined on it that are similar to the stochastic process. In particular, such a structure should be an *ordered distance space*. And ideally, the stochastic process and the ordered distance space should be both *isomorphic* and *isometric* with respect to each other.

This text proposes two *ordered distance spaces* for stochastic processing:

**Chapter 2** proposes mapping to *directed weighted graphs*. In such a structure, order is represented by direction of it’s edges, and distance by the lengths of it’s edges. Furthermore, probability can be represented by weights assigned to it's vertices.

---

Chapter 3 proposes mapping to an ordered distance linear space \((\mathbb{R}^n, \leq, \cdot, +, \cdot, \times)\), where \((\mathbb{R}, +, \cdot)\) is a field, \(\cdot\) is the vector addition operator on \(\mathbb{R}^n \times \mathbb{R}^n\), and \(\cdot\) is the scalar-vector multiplication operator on \(\mathbb{R} \times \mathbb{R}^n\). Probability has no natural representation here, and must be assigned through a separate probability function \(P\).

Mapping to a weighted graph as in Chapter 2 is useful for analysis of a single random variable \(X\). The traditional expectation value \(E(X)\) of \(X\) is often a poor choice of a statistic. For example, the traditional expected value of a fair die is \(E(X) = \frac{1}{6}(1 + 2 + \cdots + 6) = 3.5\). But this result has no relationship with reality or with intuition because the result implies that we expect the value of \(\Box\) or \(\Diamond\) more than we expect the outcome of say \(\Box\) or \(\Box\). The fact is, that for a fair die, we would expect any pair of values equally. The reason for this is that the values of the face of a fair die are merely symbols with no order, and with no metric geometry other than the discrete metric geometry. Weighted graphs on the other hand offer structures more similar to the underlying stochastic process. And the expectation \(E(X)\) of \(X\) can be defined simply as the center of the weighted graph (as illustrated above for a weighted die with 3 shaded in blue).

However, the mapping has limitations with regards to a sequence of random variables in performing sequence analysis (using for example Fourier analysis or wavelet analysis), in performing sequence processing (using for example FIR filtering or IIR filtering), in making diagnostic measurements (using a post-transform metric space), or in making “optimal” decisions (based on “distance” measurements in a metric space or more generally a distance space).

Mapping to an ordered distance linear space \(Y\) as in Chapter 3 provides the linear space component of \(Y\), which in turn provides a much more convenient framework for sequence analysis and processing (as compared to the weighted graph). The ordered set and distance space components of \(Y\) allow one to preserve the order structure and distance geometry inherent in the underlying stochastic process, which can provide a less distorted (as compared to the real line) framework for analysis, diagnostics, and optimal decision making.

For any given stochastic process, there are an infinite number of possible random variable mappings, with some being “better”—with respect to some criteria (probably involving isomorphic and isometric properties)—than others. Multiple mappings using random variables \(W, X, Y, \) and \(Z\) for the “real die” stochastic process \(G\) are illustrated in Figure 1.1 (page 4).

1.2 Further mathematical support

Further mathematical support for these two methods is provided in appendices. Here is partial summary of what is found there:
Figure 1.1: several random variable mappings for the real die

**APPENDIX A** introduces what is herein called the Lagrange arc distance function. It is important in this text because it is used in Chapter 3 in the processing of real die sequences in $\mathbb{R}^3$ and spinner sequences in $\mathbb{R}^2$. The function is an extension to all of $\mathbb{R}^N$ of the spherical metric, which has as domain only a “spherical” surface in $\mathbb{R}^N$. The Lagrange arc distance $d(p, q)$ draws arcs between certain pairs of points $(p, q)$ using Lagrange interpolation.

**APPENDIX B** presents distance spaces. A distance space is a metric space in which the triangle inequality does not necessarily hold. Distance spaces are important in this text because the Lagrange arc distance is a distance function, and not a metric in that it is not always true that $d(p, r) \leq d(p, q) + d(q, r)$.

**APPENDIX C** introduces what is herein called the power distance space. This space is a generalization of the metric space, and is a distance space that satisfies what is herein called the power triangle inequality:

$$d(x, y) \leq 2\sigma \left[ \frac{1}{2}d^p(x, z) + \frac{1}{2}d^p(z, y) \right]^{\frac{1}{2}} \quad \forall (p, x) \in \mathbb{R}^+ \times \mathbb{R}^+.$$ 

It turns out that the inequality $2\sigma \leq 2^{\frac{1}{n}}$ has special significance with regards to these spaces and appears repeatedly in Appendix C. It is plotted in Figure 1.2 (page 5). Power distance spaces are not explicitly used in this text. However, they may prove useful to future research in symbolic sequence processing in which the triangle inequality fails to hold, but in which some of the key properties of a metric space are still required.

### 1.3 Standard definitions

#### 1.3.1 Sets

**Definition 1.1.** Let $X$ be a set. The empty set $\emptyset$ is defined as $\emptyset \triangleq \{ x \in X | x \neq x \}$.

---

3 Halmos (1960) page 8, Kelley (1955) page 3, Kuratowski (1961), page 26
CHAPTER 1. FOUNDATIONS

geometric inequality at \((p, \sigma) = (0, \frac{1}{2})\)

2nd order inflection point at \((-\frac{1}{2} \ln 2, \frac{1}{2} e^{-2})\)

harmonic inequality at \((p, \sigma) = (-1, \frac{1}{2})\)

to minimal inequality at \((p, \sigma) = (\infty, \frac{1}{2})\)

to inframetric inequality at \((p, \sigma) = (\infty, \frac{1}{2})\)

\[ \sigma = \frac{1}{2}(2^\frac{1}{2}) = 2^{\frac{1}{2} - 1} \text{ or } p = \frac{\ln^2}{\ln(2\rho)} \]

Figure 1.2: \( \sigma = \frac{1}{2}(2^\frac{1}{2}) = 2^{\frac{1}{2} - 1} \) or \( p = \frac{\ln^2}{\ln(2\rho)} \)

**Definition 1.2.** Let \( \mathbb{R} \) be the set of real numbers. Let \( \mathbb{R}^+ \triangleq \{ x \in \mathbb{R} \mid x \geq 0 \} \) be the set of non-negative real numbers. Let \( \mathbb{R}^+ \triangleq \{ x \in \mathbb{R} \mid x > 0 \} \) be the set of positive real numbers. Let \( \mathbb{R}^* \triangleq \mathbb{R} \cup (-\infty, \infty) \) be the set of extended real numbers. Let \( \mathbb{Z} \) be the set of integers. Let \( \mathbb{W} \triangleq \{ n \in \mathbb{Z} \mid n \geq 0 \} \) be the set of whole numbers. Let \( \mathbb{N} \triangleq \{ n \in \mathbb{Z} \mid n \geq 1 \} \) be the set of natural numbers. Let \( \mathbb{Z}^+ \triangleq \mathbb{Z} \cup (-\infty, \infty) \) be the extended set of integers.

### 1.3.2 Relations

One of the most fundamental structures in mathematics is the ordered pair, and one of the most common definitions of ordered pair is due to Kuratowski (1921) and is presented next:

**Definition 1.3.** The ordered pair \((a, b)\) is defined as \((a, b) \triangleq \{ \{a\}, \{a, b\}\}\).

**Proposition 1.1** (next) and **Corollary 1.1** demonstrate that the definition of ordered pair given by **Definition 1.3** allows \(a\) and \(b\) to be unambiguously extracted from \((a, b)\) and that \((a, b)\) is well defined.

**Proposition 1.1.**

\[
\begin{align*}
\{a\} &= \bigcap (a, b) = \bigcap \{\{a\}, \{a, b\}\} = \{a\} \cap \{a, b\} \\
\{b\} &= \bigtriangleup (a, b) = \bigtriangleup \{\{a\}, \{a, b\}\} = \{a\} \bigtriangleup \{a, b\}
\end{align*}
\]

**Corollary 1.1.** \((a, b) = (c, d) \iff \{a = c \text{ and } b = d\}\)

\[
\begin{align*}
\{a\} &= \bigcap (a, b) = \bigcap (c, d) = \{c\} & \text{by Proposition 1.1 and left hypothesis} \\
\{b\} &= \bigtriangleup (a, b) = \bigtriangleup (c, d) = \{d\} & \text{by Proposition 1.1 and left hypothesis} \\
(a, b) &= (c, d) & \text{by right hypothesis}
\end{align*}
\]

---

4The mathematical structure called set is left undefined in this paper. For more information on sets, see for example Zermelo (1908a) pages 263–267 (7 axioms), Zermelo (1908b) (English translation of previous reference), Fraenkel (1922), Halmos (1960) pages 1–6 (Naive set theory), Wolf (1998), page 139

5Notation \(\mathbb{W}, \mathbb{N}, \text{ etc.: Bourbaki notation. References: } \text{ Davis (2005) page 9, Cohn (2012) page 3}


7As an alternative to the Kuratowski definition, the ordered pair can also be taken as an axiom. References: Bourbaki (1968), page 72, Munkres (2000), page 13


9Apostol (1975) page 33, Hausdorff (1937) page 15

Daniel J. Greenhoe
Definition 1.4. Let $X$ and $Y$ be sets. The Cartesian product $X \times Y$ is defined as 
$$X \times Y \triangleq \{(x, y) \mid (x \in X) \text{ and } (y \in Y)\}$$

Definition 1.5. Let $X$ and $Y$ be sets. A relation $\mathcal{R}$ on $X$ and $Y$ is any subset of $X \times Y$ such that $\mathcal{R} \subseteq X \times Y$. The set $2^{XY}$ is the set of all relations in $X \times Y$.

Definition 1.6. Let $X$ and $Y$ be sets. A relation $f \in 2^{XY}$ is a function if 
$$\{(x, y_1) \in f \text{ and } (x, y_2) \in f\} \implies \{y_1 = y_2\}.$$ 
The set $Y^X$ is the set of all functions in $2^{XY}$.

A function does not always have an inverse that is also a function. But unlike functions, every relation has an inverse that is also a relation. Note that since all functions are relations, every function does have an inverse that is at least a relation, and in some cases this inverse is also a function.

Definition 1.7. Let $\mathcal{R}$ be a relation in $2^{XY}$. $\mathcal{R}^{-1}$ is the inverse of relation $\mathcal{R}$ if 
$$\mathcal{R}^{-1} \triangleq \{(y, x) \in Y \times X \mid (x, y) \in \mathcal{R}\}.$$ 
The inverse relation $\mathcal{R}^{-1}$ is also called the converse of $\mathcal{R}$.

Definition 1.8. Let $X$ be a set. The quantity $2^X$ is the power set of $X$ such that 
$$2^X \triangleq \{A \subseteq X\}$$ (the set of all subsets of $X$).

Definition 1.9. Let $Y$ be a set. The structure $Y^n$ for $n \in \mathbb{N}$ is a set defined as 
$$Y^1 \triangleq Y$$ 
$$Y^n \triangleq Y \times Y^{n-1} \text{ for } n = 2, 3, 4, \ldots$$

Definition 1.10. The set of complex numbers $\mathbb{C}$ is defined as $\mathbb{C} \triangleq \mathbb{R}^2$.

Definition 1.11. Let $Y_1, Y_2, \ldots, Y_n$ be sets. The structure $(x_1, x_2, \ldots, x_n)$ is an $n$-tuple on $Y_1 \times Y_2 \times \cdots \times Y_n$ if $(x_1, x_2, \ldots, x_n)$ is an element in the set $Y_1 \times Y_2 \times \cdots \times Y_n$. A 3-tuple is also called an ordered triple or simply a triple.

Definition 1.12. Let $\mathcal{R} \in 2^{XY}$ be a relation (Definition 1.5 page 6).

The domain of $\mathcal{R}$ is $D(\mathcal{R}) \triangleq \{x \in X \mid \exists y \text{ such that } (x, y) \in \mathcal{R}\}$.

The image set of $\mathcal{R}$ is $I(\mathcal{R}) \triangleq \{y \in Y \mid \exists x \text{ such that } (x, y) \in \mathcal{R}\}$.

The null space of $\mathcal{R}$ is $N(\mathcal{R}) \triangleq \{x \in X \mid (x, 0) \in \mathcal{R}\}$.

The range of $\mathcal{R}$ is any set $R(\mathcal{R})$ such that $I(\mathcal{R}) \subseteq R(\mathcal{R})$.

Definition 1.13. The set function $|A| \in \mathbb{Z}^{\times 2^X}$ is the cardinality of $A$ such that 
$$|A| \triangleq \begin{cases} \text{the number of elements in } A & \text{for finite } A \\ \infty & \text{otherwise} \end{cases}$$ 
$\forall A \in 2^X$.

---

11 Maddux (2006) page 4, Halmos (1960) pages 26–30, Suppes (1972) page 86, Kelley (1955) page 10, Bourbaki (1939), Bottazzini (1986) page 7, Comtet (1974) page 4 ($\mid Y^X \mid$); The notation $2^{XY}$ is motivated by the fact that for finite $X$ and $Y$, $\mid 2^{XY} \mid = \lvert 2^X \rvert \lvert Y \rvert$.
12 Maddux (2006) page 4, Halmos (1960) pages 26–30, Suppes (1972) page 86, Kelley (1955) page 10, Bourbaki (1939), Bottazzini (1986) page 7, Comtet (1974) page 4 ($\mid Y^X \mid$); The notation $Y^X$ is motivated by the fact that for finite $X$ and $Y$, $\mid Y^X \mid = \lvert Y \rvert^{\lvert X \rvert}$.
13 Suppes (1972) page 61 (Definition 6, inverse=“converse”)
14 Kelley (1955) page 7
15 Peirce (1883) page 188 (inverse=“converse”)
16 Munkres (2000), page 16, Kelley (1955) page 7
Definition 1.14. Let $f$ be a function in $y^x$ (Definition 1.6 page 6).

- $f$ is surjective (also called onto) if $f(x) = y$.
- $f$ is injective (also called one-to-one) if $f(x) = f(y) \implies x = y$.
- $f$ is bijective (also called one-to-one and onto) if $f$ is both surjective and injective.

1.3.3 Field operator pairs

Definition 1.15. Let $X$, $Y$, and $Z$ be sets. Let $(\mathbb{R}, +, \cdot)$ be the standard field of real numbers. The addition operator $\oplus : X \times Y \rightarrow Z$ is defined as shown in Table 1.1 (page 7). Moreover, for some sequence $(x_n)_n$,

\[
\bigoplus_{n=1}^N x_n \triangleq x_1 \oplus x_2 \oplus \cdots \oplus x_N \quad \text{and} \quad \bigoplus_{n \in \mathbb{D}} x_n \triangleq x_{a \in \mathbb{D}} \oplus x_{b \in \mathbb{D}} \oplus \cdots \oplus x_{r \in \mathbb{D}}
\]

1. If $X \times Y \triangleq \mathbb{R} \times \mathbb{R}$ then $x \oplus y \triangleq x + y \in \mathbb{R}$.
2. If $X \times Y \triangleq \mathbb{C} \times \mathbb{C}$ then $(a, b) \oplus (c, d) \triangleq (a + c, b + d) \in \mathbb{C}$.
3. If $X \times Y \triangleq \mathbb{R}^n \times \mathbb{R}^n$ then $(x_1, x_2, \ldots, x_n) \oplus (y_1, y_2, \ldots, y_n) \triangleq (x_1 + y_1, \ldots, x_n + y_n) \in \mathbb{R}^n$.
4. If $X \times Y \triangleq \mathbb{C}^n \times \mathbb{C}^n$ then $(x_1, x_2, \ldots, x_n) \oplus (y_1, y_2, \ldots, y_n) \triangleq (x_1 + y_1, \ldots, x_n + y_n) \in \mathbb{C}^n$.
5. If $X \times Y \triangleq \mathbb{R} \times \mathbb{C}$ then $x \oplus (a, b) \triangleq (x + a, x + b) \in \mathbb{C}$.
6. If $X \times Y \triangleq \mathbb{C} \times \mathbb{R}$ then $x \oplus y \triangleq y \oplus x \in \mathbb{C}$.

Table 1.1: Definition of the addition operator $\oplus : X \times Y \rightarrow Z$ (see Definition 1.15 page 7)

Definition 1.16. Let $X$, $Y$, and $Z$ be sets. Let $(\mathbb{R}, +, \cdot)$ be the standard field of real numbers. Let the juxtaposition operator $xy$ on $x$ and $y$ be equivalent to the real field operator $x \cdot y$ on $x$ and $y$ such that $xy \triangleq x \cdot y$. The multiplication operator $\otimes : X \times Y \rightarrow Z$ is defined as shown in Table 1.2 (page 8).

1.3.4 Graph Theory

Definition 1.17. Let $2^{XX}$ be the set of all relations (Definition 1.5 page 6) on a set $X$. The pair $(X, E)$ is a graph if $E \subseteq 2^{XX}$. A graph $(X, E)$ is undirected if $(x, y) \in E \implies (y, x) \in E$. A graph $(X, E)$ is directed if it is not undirected. A graph that is directed is an undirected graph. The elements of $X$ are the vertices and the ordered pairs of $E$ are the arcs of a graph $(X, E)$. The element $x$ is the tail and $y$ is the head of an arc $(x, y)$.

17orgen Bang-Jensen and Gutin (2007), page 2 (§1.2), Haray (1969), page 9
1. If $X \times Y \triangleq \mathbb{R} \times \mathbb{R}$ then $x \otimes y \triangleq xy \in \mathbb{R}$.
2. If $X \times Y \triangleq \mathbb{R} \times \mathbb{C}$ then $x \otimes (a, b) \triangleq (xa, xb) \in \mathbb{C}$.
3. If $X \times Y \triangleq \mathbb{C} \times \mathbb{C}$ then $(a, b) \otimes (c, d) \triangleq (ac - bd, ad + bc) \in \mathbb{C}$.
4. If $X \times Y \triangleq \mathbb{R} \times \mathbb{R}^n$ then $x \otimes (y_1, y_2, \ldots, y_n) \triangleq (xy_1, xy_2, \ldots, xy_n) \in \mathbb{R}^n$.
5. If $X \times Y \triangleq \mathbb{C} \times \mathbb{R}^n$ then $x \otimes (y_1, y_2, \ldots, y_n) \triangleq (y_1 \otimes x, y_2 \otimes x, \ldots, y_n \otimes x) \in \mathbb{C}^n$.
6. If $X \times Y \triangleq \mathbb{C} \times \mathbb{C}^n$ then $x \otimes (y_1, y_2, \ldots, y_n) \triangleq (x \otimes y_1, x \otimes y_2, \ldots, x \otimes y_n) \in \mathbb{C}^n$.
7. If $X \times Y \triangleq \mathbb{R}^n \times \mathbb{R}$ then $x \otimes y \triangleq y \otimes x \in \mathbb{R}^n$.
8. If $X \times Y \triangleq \mathbb{C} \times \mathbb{C}^n$ then $x \otimes y \triangleq y \otimes x \in \mathbb{C}^n$.
9. If $X \times Y \triangleq \mathbb{C} \times \mathbb{C}^n$ then $x \otimes y \triangleq y \otimes x \in \mathbb{C}^n$.

Table 1.2: Definition of the multiplication operation $\otimes : X \times Y \rightarrow Z$ (see Definition 1.16 page 7)

Definition 1.18. The tuple $(X, E, d, w)$ is a weighted graph if $(X, E)$ is a graph (Definition 1.17 page 7), $d(x, y)$ is a function in $(\mathbb{R}^+)^{X \times X}$ (Definition 1.6 page 6), and $w(x)$ is a function in $(\mathbb{R}^+)^{X}$. The function $d$ is called the edge weight, and the function $w$ is called the vertex weight.

Definition 1.19. Let $G \triangleq (X, E, d, w)$ be a weighted graph (Definition 1.18 page 8). The center $\bar{G}(G)$ of $G$ is $\bar{G}(G) \triangleq \text{arg min } \max_{x \in X} d(x, y) w(y)$.

1.4 Traditional probability

"While writing my book I had an argument with Teller. He asserted that everyone said “random variable” and I asserted that everyone said “chance variable.” We obviously had to use the same name in our books, so we decided the issue by a stochastic procedure. That is, we tossed for it and he won."

Joseph Leonard Doob (1910–2004), pioneer of and key contributor to mathematical probability

Definition 1.20. 19 Let $X$ be a set. A function $P \in \mathbb{R}^+^X$ is a probability function if

1. $P(1) = 1$ (NORMALIZED) and
2. $P(x) \geq 0 \forall x \in X$ (NONNEGATIVE) and
3. $x \land y = 0 \implies P(x \lor y) = P(x) + P(y) \forall x, y \in X$ (ADDITIONAL).

Definition 1.21. 20 The triple $(\Omega, \mathcal{E}, P)$ is a probability space if $\Omega$ is a set, $\mathcal{E}$ is a $\sigma$-algebra on $\Omega$, and $P$ is a probability function in $[0, 1]^\mathcal{E}$. In this case, $\Omega$ is called the set of outcomes.

Before defining a random variable formally, note two things that a random variable is not:21

- A random variable is not random.
- A random variable is not a variable.

What is it then? It is a function (next definition). In particular, it is a function that maps from an underlying stochastic process into $\mathbb{R}$. Any “randomness” (whatever that means) it may appear to

---

20 \(\text{Greenhoe (2015)}\)
have comes from the stochastic process it is mapping from. But the function itself (the random
variable itself) is very deterministic and well-defined.

**Definition 1.22.** A traditional random variable \( X \) on a probability space \((\Omega, \mathcal{E}, P)\) is
any function in the set \( \mathbb{R}^\Omega \) (Definition 1.6 page 6).

**Definition 1.23.** Let \((\Omega, \mathcal{E}, P)\) be a probability space (Definition 1.21 page 8) and \( X \in \mathbb{R}^\Omega \) a random vari-
able (Definition 1.22 page 9).

- The traditional expected value \( E(X) \) of \( X \) is \( E(X) \Delta \int_{\mathbb{R}} x p(x) \, dx \).
- The traditional variance \( \text{Var}(X) \) of \( X \) is \( \text{Var}(X) \Delta \int_{\mathbb{R}} [x - E(X)]^2 p(x) \, dx \).

**Proposition 1.2.** Let \((\Omega, \mathcal{E}, P)\) be a probability space (Definition 1.21 page 8), \( X \in \mathbb{R}^\Omega \) a traditional random variable (Definition 1.22 page 9), \( E(X) \) the traditional expected value of \( X \), and \( \text{Var}(X) \) the traditional variance of \( X \) (Definition 1.23 page 9).

\[
\{ P(x) = 0 \ \forall x \notin \mathbb{Z} \} \implies \begin{cases} 1. \ E(X) = \sum_{x \in \mathbb{Z}} xP(x) & \text{and} \\ 2. \ \text{Var}(X) = \sum_{x \in \mathbb{Z}} [x - E(X)]^2 P(x) \end{cases}
\]

**Proposition 1.3.** Let \((\Omega, \mathcal{E}, P), X, \) and \( E \) be defined as in Proposition 1.2 (page 9).

\[ P(y-x) = P(y+x) \ \forall x \in \mathbb{R} \implies \{ E(X) = y \}\]

\[\text{PROOF:}\]

\[
\begin{align*}
E(X) &= \int_{-\infty}^{\infty} x p(x) \, dx \\
&= \int_{-\infty}^{\gamma} x p(x) \, dx + \int_{\gamma}^{\infty} x p(x) \, dx \\
&= \int_{-\infty}^{u+\gamma} (u+\gamma)p(u+\gamma) \, du + \int_{u+\gamma}^{\infty} (u+\gamma)p(u+\gamma) \, du \quad \text{where } u \triangleq x - \gamma \implies x = u + \gamma \\
&= \int_{-\infty}^{\infty} (u+\gamma)p(u+\gamma) \, du \\
&= \gamma \left( \int_{-\infty}^{0} p(w+\gamma) \, dw + \int_{0}^{\infty} p(u+\gamma) \, du \right) \\
&= \gamma \left( \int_{-\infty}^{\infty} p(u+\gamma) \, du \right)
\end{align*}
\]

\[\quad \text{by definition of } E \ (\text{Definition 1.23 page 9) \ and \ additive \ property \ of \ Lebesgue \ integration \ op.}^{23}\]

\[
\begin{align*}
\text{Var}(X) &= \int_{-\infty}^{\infty} [x - E(X)]^2 p(x) \\
&= \int_{-\infty}^{\infty} [x - (u+\gamma)]^2 p(u+\gamma) \, du \\
&= \int_{-\infty}^{\infty} [x - ((u+\gamma)+\gamma)]^2 p((u+\gamma)+\gamma) \, du \\
&= \int_{-\infty}^{\infty} [x - (u+\gamma+\gamma)]^2 p(u+\gamma+\gamma) \, du \\
&= \int_{-\infty}^{\infty} [x - ((u+\gamma)+\gamma)]^2 p((u+\gamma)+\gamma) \, du \\
&= \int_{-\infty}^{\infty} [x - (u+\gamma+\gamma)]^2 p(u+\gamma+\gamma) \, du \\
&= \int_{-\infty}^{\infty} [x - (u+\gamma+\gamma)]^2 p(u+\gamma+\gamma) \, du \\
&= \int_{-\infty}^{\infty} [x - (u+\gamma+\gamma)]^2 p(u+\gamma+\gamma) \, du \\
&= \gamma \left( \int_{-\infty}^{0} p(w+\gamma) \, dw + \int_{0}^{\infty} p(u+\gamma) \, du \right) \\
&= \gamma \left( \int_{-\infty}^{\infty} p(u+\gamma) \, du \right)
\end{align*}
\]

\[\quad \text{by symmetry hypothesis; cancels to } 0\]

\[\gamma \left( \int_{-\infty}^{0} p(u+\gamma) \, du \right) = \gamma \left( \int_{-\infty}^{\infty} p(u+\gamma) \, du \right) = \gamma \int_{-\infty}^{\infty} p(u) \, du = \gamma \]

\[\]
1.5 Order space concepts

1.5.1 Order

Definition 1.24. 24 Let $X$ be a set. A relation $\leq$ is an **order relation** in $2^X$ (Definition 1.5 page 6) if
1. $x \leq x$ $\forall x \in X$ (reflexive) and
2. $x \leq y$ and $y \leq z$ $\implies$ $x \leq z$ $\forall x, y, z \in X$ (transitive) and
3. $x \leq y$ and $y \leq x$ $\implies$ $x = y$ $\forall x, y \in X$ (anti-symmetric).

An **ordered set** is the pair $(X, \leq)$. If $\leq \subseteq \emptyset$ (Definition 1.1 page 4), then $(X, \leq)$ is an **unordered set**. If $x \leq y$ or $y \leq x$, then elements $x$ and $y$ are said to be **comparable**; otherwise they are **incomparable**.

Definition 1.25. 25 A relation $\leq$ is a **linear order relation** on $X$ if
1. $\leq$ is an **ORDER RELATION** (Definition 1.24 page 10) and
2. $x \leq y$ or $y \leq x$ $\forall x, y \in X$ (comparable).

A **linearly ordered set** is the pair $(X, \leq)$. A linearly ordered set is also called a **totally ordered set**, a **fully ordered set**, and a **chain**.

The familiar relations $\geq$, $<$, and $>$ (next) can be defined in terms of the order relation $\leq$ (Definition 1.24—previous).

Definition 1.26. 26 Let $(X, \leq)$ be an ordered set. The relations $\geq$, $<$, $\in X^2$ are defined as follows:

\[ x \geq y \iff y \leq x \qquad \forall x, y \in X \]

\[ x \leq y \iff x \leq y \text{ and } x \neq y \qquad \forall x, y \in X \]

\[ x \geq y \iff x \geq y \text{ and } x \neq y \qquad \forall x, y \in X \]

The relation $\geq$ is called the **dual** of $\leq$.

**Example 1.1** (Coordinatewise order relation). 27 Let $(X, \leq)$ be an ordered set.
Let $x \triangleq (x_1, x_2, \ldots, x_n)$ and $y \triangleq (y_1, y_2, \ldots, y_n)$.

The **coordinatewise order relation** $\leq$ on the Cartesian product $X^n$ is defined for all $x, y \in X^n$ as

\[ x \leq y \iff \{ x_1 \leq y_1 \text{ and } x_2 \leq y_2 \text{ and } \ldots \text{ and } x_n \leq y_n \} \]

**Example 1.2** (Lexicographical order relation). 28 Let $(X, \leq)$ be an ordered set.
Let $x \triangleq (x_1, x_2, \ldots, x_n)$ and $y \triangleq (y_1, y_2, \ldots, y_n)$.

The **lexicographical order relation** $\leq$ on the Cartesian product $X^n$ is defined for all $x, y \in X^n$ as

\[ x \leq y \iff \left\{ \begin{array}{ll}
\mbox{or} & x_1 < y_1 \\
\mbox{or} & x_2 < y_2 \\
\mbox{or} & x_3 < y_3 \\
\mbox{or} & (x_1, x_2) = (y_1, y_2) \\
\mbox{or} & (x_1, x_2, \ldots, x_{n-1}) = (y_1, y_2, \ldots, y_{n-1}) \\
\mbox{or} & x_n \leq y_n \\
\end{array} \right\} \]

The lexicographical order relation is also called the **dictionary order relation** or **alphabetic order relation**.

---

24. MacLane and Birkhoff (1999) page 470, Beran (1985) page 1, Korselt (1894) page 156 (I, II, (1)), Dedekind (1900) page 373 (I–III). An order relation is also called a partial order relation. An ordered set is also called a partially ordered set or poset.
25. MacLane and Birkhoff (1999) page 470, Ore (1935) page 410
26. Peirce (1880) page 2
27. Shen and Vereshchagin (2002) page 43
Definition 1.27. 29 In an ordered set \((X, \leq)\),
the set \([x : y] \triangleq \{ z \in X | x \leq z \leq y \}\) is a closed interval on \((X, \leq)\)
and
the set \((x : y) \triangleq \{ z \in X | x < z \leq y \}\) is a half-open interval on \((X, \leq)\)
and
the set \([x : y) \triangleq \{ z \in X | x \leq z < y \}\) is a half-open interval on \((X, \leq)\)
and
the set \((x : y) \triangleq \{ z \in X | x < z < y \}\) is an open interval. on \((X, \leq)\)

Definition 1.28. Let \((\mathbb{R}, \leq)\) be the ordered set of real numbers (Definition 1.24 page 10).
The absolute value \(|\cdot| \in \mathbb{R}^\mathbb{R}\) is defined as^30
\[ |x| \triangleq \begin{cases} -x & \text{for } x \leq 0 \\ x & \text{otherwise} \end{cases} \]

Definition 1.29. 31 Let \((X, \leq)\) be an ordered set. A subset \(D \subseteq X\) is convex in \(X\) if
\[ x, y \in D \implies (x : y) \subseteq D. \]
Example 1.3. Convex subsets of \(Z\) under the usual integer ordering relation include
\(\emptyset, \mathbb{Z}, \mathbb{W}, \mathbb{N}, \{0, 1, 2, 3, 4, 5\}\), and \\{-2, -1, 0, 1, 2, 3\}\.

Definition 1.30. Let \((X, \leq)\) be an ordered set. For any set \(A \subseteq 2^X, c\) is an upper bound of \(A\) in \((X, \leq)\)
if \( x \leq c \quad \forall x \in A \). An element \(b\) is the least upper bound, or lub, of \(A\) in \((X, \leq)\) if
\[ b \text{ and } c \text{ are upper bounds of } A \implies b \leq c. \]
The least upper bound of the set \(A\) is denoted \(\bigvee A\).
The join \(x \lor y\) of \(x\) and \(y\) is defined as \(x \lor y \triangleq \bigvee \{x, y\}\).

Definition 1.31. Let \((X, \leq)\) be an ordered set. For any set \(A \subseteq 2^X, p\) is a lower bound of \(A\) in \((X, \leq)\) if
\( p \leq x \quad \forall x \in A \). An element \(a\) is the greatest lower bound, or glb, of \(A\) in \((X, \leq)\) if
\[ a \text{ and } p \text{ are lower bounds of } A \implies p \leq a. \]
The greatest lower bound of the set \(A\) is denoted \(\bigwedge A\).
The meet \(x \land y\) of \(x\) and \(y\) is defined as \(x \land y \triangleq \bigwedge \{x, y\}\).

Lemma 1.1. Let \((X, \leq)\) be an ordered set (Definition 1.24 page 10). Let \(\bigvee A\) be the least upper bound
(Definition 1.30 page 11) of a set \(A \subseteq 2^X\) (Definition 1.8 page 6). Let \(\bigwedge A\) be the greatest lower bound (Definition 1.31
page 11) of a set \(A \subseteq 2^X\).
\[ \{A = X\} \implies \begin{cases} 1. \bigvee A = \{a \in X | x \leq a, \forall x, a \in X\} \quad \text{and} \\ 2. \bigwedge A = \{a \in X | a \leq x, \forall x, a \in X\}. \end{cases} \]

Definition 1.32. Let \((\mathbb{R}, \leq)\) be the standard ordered set of real numbers. The floor function
\(|x| \in \mathbb{R}^\mathbb{R}\) and the ceiling function \(|x| \in \mathbb{Z}^\mathbb{R}\) are defined as
\[ |x| \triangleq \bigvee \{n \in \mathbb{Z} | n \leq x\} \quad \text{and} \quad |x| \triangleq \bigwedge \{n \in \mathbb{Z} | n \geq x\}. \]

### 1.5.2 Lattices

The structure available in an ordered set (Definition 1.24 page 10) tends to be insufficient to ensure “well-behaved” mathematical systems. This situation is greatly remedied if every pair of elements in the

---

29 Apostol (1975) page 4, Ore (1935) page 409, Duthie (1942) page 2, Ore (1935) page 425 (quotient structures)
30 A more general definition for absolute value is available for any commutative ring: Let \(R\) be a commutative ring.
A function \(|\cdot|\) in \(R^R\) is an absolute value, or modulus, on \(R\) if
1. \(|x| \geq 0 \quad x \in R\) (non-negative) and
2. \(|x| = 0 \iff x = 0 \quad x \in R\) (nondegenerate) and
3. \(|xy| = |x| \cdot |y| \quad x, y \in R\) (homogeneous / submultiplicative) and
4. \(|x + y| \leq |x| + |y| \quad x, y \in R\) (subadditive / triangle inequality)

Reference: Cohn (2002) page 312
31 Barvinok (2002) page 5
ordered set has both a least upper bound and a greatest lower bound (Definition 1.31 page 11) in the set; in this case, that ordered set is a lattice (next definition).\(^\text{32}\)

**Definition 1.33.**\(^\text{33}\) An algebraic structure \(L = (X, \lor, \land; \leq)\) is a lattice if

1. \((X, \leq)\) is an ordered set (Definition 1.24 page 10) and
2. \(x, y \in X \implies x \lor y \in X\) (Definition 1.30 page 11) and
3. \(x, y \in X \implies x \land y \in X\) (Definition 1.31 page 11).

The lattice \(L\) is linear if \((X, \leq)\) is a linearly ordered set (Definition 1.25 page 10).

**Theorem 1.1.**\(^\text{34}\) \((X, \lor, \land; \leq)\) is a lattice (Definition 1.33 page 12) if

\[
\begin{align*}
\lor: & \quad x \lor x = x \\
\land: & \quad x \land x = x \\
(x \lor y) \lor z = x \lor (y \lor z) \\
x \lor (x \land y) = x
\end{align*}
\]

\(\lor: \quad x \lor y = y \lor x \quad (\lor\text{ is commutative})\) and

\(\land: \quad x \land y = y \land x \quad (\land\text{ is commutative})\) and

\(\lor: \quad (x \lor y) \land z = x \land (y \lor z) \quad (\lor\text{ is associative})\) and

\(\land: \quad x \land (x \lor y) = x \quad (\land\text{ is associative})\) and

\(\forall x \in X \quad (\text{idempotent})\) and

\(\forall x, y \in X \quad (\text{commutative})\) and

\(\forall x, y, z \in X \quad (\text{associative})\) and

\(\forall x, y \in X \quad (\text{absorptive}).\)

**Minimax inequality.** Suppose we arrange a finite sequence of values into \(m\) groups of \(n\) elements per group. This could be represented as an \(m \times n\) matrix. Suppose now we find the minimum value in each row, and the maximum value in each column. We can call the maximum of all the minimum row values the *maximin*, and the minimum of all the maximum column values the *minimax*. Now, which is greater, the maximin or the minimax? The minimax inequality demonstrates that the maximin is always less than or equal to the minimax. The minimax inequality is illustrated below and stated formerly in Theorem 1.2 (page 12).

\[
\begin{align*}
\text{maximin} & = \left\lfloor \bigwedge_{i=1}^{m} \left\lfloor \bigvee_{j=1}^{n} x_{ij} \right\rfloor \right\rfloor \\
\text{minimax} & = \left\lceil \bigvee_{i=1}^{m} \left\lfloor \bigwedge_{j=1}^{n} x_{ij} \right\rfloor \right\rceil
\end{align*}
\]

**Theorem 1.2 (minimax inequality).**\(^\text{35}\) Let \((X, \lor, \land; \leq)\) be a lattice (Definition 1.33 page 12).

\[
\bigvee_{i=1}^{m} \bigwedge_{j=1}^{n} x_{ij} \leq \bigwedge_{i=1}^{m} \bigvee_{j=1}^{n} x_{ij} \quad \forall x_{ij} \in X
\]

*maximin: largest of the smallest

*minimax: smallest of the largest*

---

\(^{32}\) Gian-Carlo Rota (1932–1999) has illustrated the advantage of lattices over simple ordered sets by pointing out that the ordered set of partitions of an integer “is fraught with pathological properties”, while the lattice of partitions of a set “remains to this day rich in pleasant surprises”. \(\text{Rota (1997) page 1440 (illustration)}, \text{ Rota (1964) page 498 (illustrations of a set)}\)

\(^{33}\) MacLane and Birkhoff (1999) page 473, Birkhoff (1948) page 16, Ore (1935), Birkhoff (1933) page 442, Maeda and Maeda (1970), page 1

\(^{34}\) MacLane and Birkhoff (1999) pages 473–475 (Lemma 1, Theorem 4), Burris and Sankappanavar (1981) pages 4–7, Birkhoff (1938), pages 795–796, Ore (1935) page 409 (a), Birkhoff (1933) page 442, Dedekind (1900) pages 371–372 (1(1)–4)

\(^{35}\) Birkhoff (1948) pages 19–20
\textbf{Proof:}

\[
\left( \bigwedge_{k=1}^{n} x_{ik} \right) \leq x_{ij} \leq \left( \bigvee_{k=1}^{n} x_{kj} \right) \quad \forall i, j
\]

smallest for any given \( i \), largest for any given \( j \)

\[\implies\]

\[
\left( \bigvee_{i=1}^{m} \left( \bigwedge_{k=1}^{n} x_{ik} \right) \right) \leq \left( \bigwedge_{j=1}^{n} \left( \bigvee_{k=1}^{m} x_{kj} \right) \right)
\]

largest among all \( i \)'s of the smallest values

\[
\implies \left( \bigvee_{i=1}^{m} \left( \bigwedge_{j=1}^{n} x_{ij} \right) \right) \leq \left( \bigwedge_{j=1}^{n} \left( \bigvee_{i=1}^{m} x_{ij} \right) \right) \quad \text{(change of variables)}
\]

\[\implies\]

Special cases of the minimax inequality include three distributive inequalities (next theorem). If for some lattice any one of these inequalities is an equality, then all three are equalities; and in this case, the lattice is a called a \textit{distributive} lattice.

\textbf{Theorem 1.3} (distributive inequalities). \footnote{Davey and Priestley (2002) page 85, Grätzer (2003) page 38, Birkhoff (1933) page 444, Korselt (1894) page 157, Müller-Olm (1997) page 13 (terminology)} \( (X, \lor, \land; \leq) \) is a lattice (Definition 1.33 page 12) \[\implies\]

\[
\begin{aligned}
&x \land (y \lor z) \geq (x \land y) \lor (x \land z) &\forall x, y, z \in X \quad \text{(join super-distributive) and} \\
x \lor (y \land z) \leq (x \lor y) \land (x \lor z) &\forall x, y, z \in X \quad \text{(meet sub-distributive) and} \\
(x \land y) \lor (x \land z) \lor (y \land z) \leq (x \lor y) \land (x \lor z) \land (y \lor z) &\forall x, y, z \in X \quad \text{(median inequality).}
\end{aligned}
\]

\[\implies\]

\textbf{Proof:}

1. Proof that \( \land \) sub-distributes over \( \lor \):

\[
(x \land y) \lor (x \land z) \leq (x \lor y) \land (y \lor z) \quad \text{by minimax inequality (Theorem 1.2 page 12)}
\]

\[
= x \land (y \lor z) \quad \text{by idempotent property of lattices (Theorem 1.1 page 12)}
\]

\[
\left\{ \bigwedge_{y} \right\} \left\{ \bigvee_{x} \right\} \left\{ \bigvee_{z} \right\} \leq \left\{ \bigvee_{y} \right\} \left\{ \bigwedge_{x} \right\} \left\{ \bigwedge_{z} \right\}
\]

2. Proof that \( \lor \) super-distributes over \( \land \):

\[
x \lor (y \land z) = (x \lor x) \lor (y \land z) \quad \text{by idempotent property of lattices (Theorem 1.1 page 12)}
\]

\[
\leq (x \lor y) \land (x \lor z) \quad \text{by minimax inequality (Theorem 1.2 page 12)}
\]

\[
\left\{ \bigwedge_{x} \right\} \left\{ \bigvee_{y} \right\} \left\{ \bigvee_{z} \right\} \leq \left\{ \bigvee_{x} \right\} \left\{ \bigwedge_{y} \right\} \left\{ \bigwedge_{z} \right\}
\]

3. Proof that of median inequality: by minimax inequality (Theorem 1.2 page 12)

\[\implies\]

Besides the distributive property, another consequence of the minimax inequality is the \textit{modularity inequality} (next theorem). A lattice in which this inequality becomes equality is said to be \textit{modular}.
Theorem 1.4 (Modular inequality). \(37\) Let \((X, \lor, \land; \leq)\) be a LATTICE (Definition 1.33 page 12).

\[
x \leq y \implies x \lor (y \land z) \leq y \land (x \lor z)
\]

\(\Box\) \begin{proof}
\[
x \lor (y \land z) = (x \land x) \lor (y \land z) \quad \text{by absorptive property (Theorem 1.1 page 12)}
\]
\[
\leq (x \lor y) \land (x \lor z) \quad \text{by the minimax inequality (Theorem 1.2 page 12)}
\]
\[
= y \land (x \lor z) \quad \text{by left hypothesis}
\]
\[
\bigwedge \left\{ \begin{array}{l}
  x \\
  y
\end{array} \right\} \bigvee \left\{ \begin{array}{l}
  z \\
  y
\end{array} \right\} \leq \bigwedge \left\{ \begin{array}{l}
  x \\
  y
\end{array} \right\} \bigvee \left\{ \begin{array}{l}
  x \\
  z
\end{array} \right\}
\]
\(\Box\)

1.5.3 Isomorphic spaces

Definition 1.34. \(38\) Let \(X \triangleq (X, \leq)\) and \(Y \triangleq (Y, \geq)\) be ordered sets.
A function \(\theta \in Y^X\) is \textit{order preserving} in \((X, Y)\) if

\[
(x \leq y) \implies (\theta(x) \preceq \theta(y)) \quad \forall x, y \in X.
\]

Definition 1.35. Let \(L_1 \triangleq (X, \lor, \land; \leq)\) and \(L_2 \triangleq (Y, \otimes, \odot; \preceq)\) be LATTICES.
\(L_1\) and \(L_2\) are \textit{isomorphic} on \((X, Y)\) if there exists a function \(\theta \in Y^X\) such that

1. \(\theta(x \lor y) = \theta(x) \otimes \theta(y) \quad \forall x, y \in X\) (preserves joins)
2. \(\theta(x \land y) = \theta(x) \odot \theta(y) \quad \forall x, y \in X\) (preserves meets).

In this case, the function \(\theta\) is said to be an \textit{isomorphism} from \(L_1\) to \(L_2\), and the isomorphic relationship between \(L_1\) and \(L_2\) is denoted as \(L_1 \cong L_2\).

Theorem 1.5. \(39\) Let \((X, \lor, \land; \leq)\) and \((Y, \otimes, \odot; \preceq)\) be lattices and \(\theta \in Y^X\) be a BIJECTIVE function with inverse \(\theta^{-1} \in X^Y\).

\[
\begin{aligned}
x_1 \leq x_2 & \implies \theta(x_1) \preceq \theta(x_2) \quad \forall x_1, x_2 \in X \quad \text{and}
\end{aligned}
\]

\[
\begin{aligned}
y_1 \preceq y_2 & \implies \theta^{-1}(y_1) \preceq \theta^{-1}(y_2) \quad \forall y_1, y_2 \in Y
\end{aligned}
\]

\(\theta\) and \(\theta^{-1}\) are order preserving

\(\iff\) \((X, \lor, \land; \leq) \cong (Y, \otimes, \odot; \preceq)\)

\(\Box\) \begin{proof}
Let \(\theta \in Y^X\) be the isomorphism between lattices \((X, \lor, \land; \leq)\) and \((Y, \otimes, \odot; \preceq)\).

1. Proof that \textit{order preserving} \(\implies\) \textit{preserves joins}:

(a) Proof that \(\theta(x_1 \lor x_2) \otimes \theta(x_1) \otimes \theta(x_2)\):

i. Note that

\[
x_1 \leq x_1 \lor x_2
\]
\[
x_2 \leq x_1 \lor x_2.
\]

ii. Because \(\theta\) is \textit{order preserving}

\[
\theta(x_1) \preceq \theta(x_1 \lor x_2)
\]
\[
\theta(x_2) \preceq \theta(x_1 \lor x_2).
\]

\(\Box\)

\(37\) Birkhoff (1948) page 19, Burris and Sankappanavar (1981) page 11, Dedekind (1900) page 374

\(38\) Burris and Sankappanavar (2000), page 10

\(39\) Burris and Sankappanavar (2000), page 10
iii. We can then finish the proof of item (1a):

\[ \theta(x_1) \otimes \theta(x_2) \leq \theta(x_1 \lor x_2) \otimes \theta(x_1 \lor x_2) \]

by order preserving hypothesis

\[ \leq \theta(x_1 \lor x_2) \]

by idempotent property page 12

(b) Proof that \( \theta(x_1 \lor x_2) \leq \theta(x_1) \otimes \theta(x_2) \):

i. Just as in item (1a), note that \( \theta^{-1}(y_1) \lor \theta^{-1}(y_2) \leq \theta^{-1}(y_1 \otimes y_2) \):

\[ \theta^{-1}(y_1) \lor \theta^{-1}(y_2) \leq \theta^{-1}(y_1 \otimes y_2) \]

by order preserving hypothesis

\[ \leq \theta^{-1}(y_1 \otimes y_2) \]

by idempotent property page 12

ii. Because \( \theta \) is order preserving

\[ \theta[\theta^{-1}(y_1) \lor \theta^{-1}(y_2)] \geq \theta(y_1 \otimes y_2) \]

by item (1b)i page 15

\[ = y_1 \otimes y_2 \]

by definition of inverse function \( \theta^{-1} \)

iii. Let \( u_1 \triangleq \theta(x_1) \) and \( u_2 \triangleq \theta(x_2) \).

iv. We can then finish the proof of item (1b):

\[ \theta(x_1 \lor x_2) = \theta[\theta^{-1}(u_1) \lor \theta^{-1}(u_2)] \]

by definition of inverse function \( \theta^{-1} \)

\[ \geq u_1 \otimes u_2 \]

by definition of \( u_1, u_2 \), item (1b)i

\[ = \theta(x_1) \otimes \theta(x_2) \]

by definition of \( u_1, u_2 \), item (1b)i

(c) And so, combining item (1a) and item (1b), we have

\[ \theta(x_1 \lor x_2) \otimes \theta(x_1) \otimes \theta(x_2) \]

and

\[ \theta(x_1 \lor x_2) \leq \theta(x_1) \otimes \theta(x_2) \]

\[ \implies \theta(x_1 \lor x_2) = \theta(x_1) \otimes \theta(x_2) \]

2. Proof that order preserving \( \implies \) preserves meets:

(a) Proof that \( \theta(x_1 \land x_2) \leq \theta(x_1) \otimes \theta(x_2) \):

\[ \theta(x_1) \otimes \theta(x_2) \otimes \theta(x_1 \land x_2) \]

by order preserving hypothesis

\[ = \theta(x_1 \land x_2) \]

by idempotent property page 12

(b) Proof that \( \theta(x_1 \land x_2) \otimes \theta(x_1) \otimes \theta(x_2) \):

i. Just as in item (2a), note that \( \theta^{-1}(y_1) \land \theta^{-1}(y_2) \geq \theta^{-1}(y_1 \otimes y_2) \):

\[ \theta^{-1}(y_1) \land \theta^{-1}(y_2) \geq \theta^{-1}(y_1 \otimes y_2) \]

by order preserving hypothesis

\[ \geq \theta^{-1}(y_1 \otimes y_2) \]

by idempotent property page 12

ii. Because \( \theta \) is order preserving

\[ \theta[\theta^{-1}(y_1) \land \theta^{-1}(y_2)] \otimes \theta(y_1 \otimes y_2) \]

by item (2b)i

\[ = y_1 \otimes y_2 \]

iii. Let \( v_1 \triangleq \theta(x_1) \) and \( v_2 \triangleq \theta(x_2) \).
iv. We can then finish the proof of item (2a):

\[
\theta(x_1 \land x_2) = \theta[\theta^{-1}(x_1) \land \theta^{-1}(x_2)] \\
= \theta[\theta^{-1}(v_1) \land \theta^{-1}(v_2)] \\
\otimes v_1 \otimes v_2 \\
= \theta(x_1) \otimes \theta(x_2)
\]

by item (2b)33

(c) And so, combining item (2a) and item (2b), we have

\[
\theta(x_1 \land x_2) \leq \theta(x_1) \otimes \theta(x_2) \quad \text{and} \\
\theta(x_1 \land x_2) \otimes \theta(x_1) \otimes \theta(x_2) \quad \text{iff} \\
\theta(x_1) = \theta(x) \otimes \theta(y)
\]

Proof that order preserving \iff isomorphic:

\[
x \leq y \implies \theta(y) = \theta(x \lor y) = \theta(x) \otimes \theta(y) \\
\implies \theta(x) \leq \theta(y) \\
x \leq y \implies \theta(x) = \theta(x \land y) = \theta(x) \otimes \theta(y) \\
\implies \theta(x) \geq \theta(y)
\]

Example 1.4. In the diagram to the right, the function \( \theta \in Y^X \) is order preserving with respect to \( \leq \) and \( \preceq \). Note that \( \theta^{-1} \) is not order preserving and that the ordered sets \( (X, \leq) \) and \( (Y, \preceq) \) are not isomorphic—as already demonstrated by Theorem 1.5 that they cannot be.

Example 1.5. In the diagram to the right, the function \( \theta \in Y^X \) is order preserving with respect to \( \leq \) and \( \preceq \). Note that \( \theta^{-1} \) is not order preserving. Like Example 1.4 (page 16), this example also illustrates the fact that that order preserving does not imply isomorphic.

Example 1.6. In the diagram to the right, the function \( \theta \in Y^X \) is order preserving with respect to \( \leq \) and \( \preceq \). Note that \( \theta^{-1} \) is also order preserving and that the ordered sets \( (X, \leq) \) and \( (Y, \preceq) \) are not isomorphic—as already demonstrated by Theorem 1.5 that they must be.

1.5.4 Monotone functions on ordered sets

Definition 1.36. Let \( (X, \leq) \) and \( (Y, \preceq) \) be ordered sets (Definition 1.24 page 10). Let \( \phi \) be a function in \( Y^X \) (Definition 1.6 page 6).

- \( \phi \) is isotone in \( Y, \preceq \) if, \( x \leq y \implies \phi(x) \preceq \phi(y) \) if \( x \leq y \).
- \( \phi \) is strictly isotone in \( Y, \preceq \) if, \( x < y \implies \phi(x) \preceq \phi(y) \) if \( x < y \).
- \( \psi \) is antitone in \( Y, \preceq \) if, \( x \leq y \implies \psi(y) \preceq \psi(x) \) if \( x \leq y \).
- \( \psi \) is strictly antitone in \( Y, \preceq \) if, \( x < y \implies \psi(y) \preceq \psi(x) \) if \( x < y \).

A function is monotone if it is isotone or antitone and strictly monotone if it is strictly isotone or strictly antitone. An isotone function in \( (Y, \preceq) \) is also said to be order preserving in \( (Y, \preceq) \).

---

40 Burris and Sankappanavar (2000), page 10
Lemma 1.2. Let $(X, \vee, \wedge; \leq)$ and $(Y, \sqcup, \sqcap; \sqsubseteq)$ be lattices \(^{42}\) (Definition 1.33 page 12). Let $f$ be a function in $X^Y$. Let $\phi$ be a function in $Y^X$ \(^{42}\) (Definition 1.6 page 6).

\[
\begin{align*}
\{ \phi \text{ is isotone} & \} \\
(\text{Definition 1.36 page 16}) \\
\implies \{ 1. \ \arg \bigoplus_{x \in X} f(x) \subseteq \arg \bigoplus_{x \in X} \phi[f(x)] \text{ and } \\
2. \ \arg \bigcap_{x \in X} f(x) \subseteq \arg \bigcap_{x \in X} \phi[f(x)] \}
\end{align*}
\]

\[\text{Proof:}\]

\[
\begin{align*}
\arg \bigoplus_{x \in X} f(x) &= \arg \bigoplus_{x \in X} \{ f(x)[f(y)] \leq f(x) \} \quad \forall x, y \in X \\
\subseteq \arg \bigoplus_{x \in X} \{ f(x)[f(y)] \leq \phi[f(x)] \} \quad \forall x, y \in X \\
= \arg \bigoplus_{x \in X} \{ f(x)[f(y)] \leq \phi[f(x)] \} \quad \forall x, y \in X \\
\triangleq \arg \bigcap_{x \in X} \phi[f(x)] \\
\end{align*}
\]

because $f \in X^X$ and by Lemma 1.1 page 11

by isotone hypothesis (Definition 1.36 page 16)

because $f \in X^X$ and by Lemma 1.1 page 11

Remark 1.1. Let $(X, \leq)$ and $(Y, \sqsubseteq)$ be ordered sets \(^{42}\) (Definition 1.24 page 10). Let $\phi$ be a function in $Y^X$ \(^{42}\) (Definition 1.6 page 6). Note that even if $\phi$ is bijective \(^{42}\) (Definition 1.14 page 7) and strictly isotone \(^{42}\) (Definition 1.36 page 16),

$x < y \iff \phi(x) \sqsubset \phi(y) \quad \forall x, y \in X$ .

An example is illustrated to the right where $\phi(l) \sqsubset \phi(r)$, but $l \not\sqsubset r$.

Lemma 1.3. Let $X \triangleq (X, \leq)$ and $Y \triangleq (Y, \sqsubseteq)$ be ordered sets. Let $\phi$ be a function in $Y^X$.

\[
\begin{align*}
\{ \begin{array}{l}
A. \ \phi \text{ is strictly isotone} \\
B. \ X \text{ and } Y \text{ are linearly ordered}
\end{array} & \} \\
\implies \{ 1. \ x \leq y \iff \phi(x) \sqsubset \phi(y) \quad \forall x, y \in X \text{ and } \\
2. \ x < y \iff \phi(x) \sqsubset \phi(y) \quad \forall x, y \in X \}
\end{align*}
\]

\[\text{Proof:}\]

\[
\begin{align*}
\phi(x) \sqsubset \phi(y) & \implies y \not\leq x \\
\implies x \leq y & \text{ by contrapositive of strictly isotone hypothesis (A)} \\
\implies \phi(x) \sqsubset \phi(y) & \text{ by linear hypothesis (B)} \\
\phi(x) \subset \phi(y) & \implies y \not\leq x \\
\implies x < y & \text{ by contrapositive of strictly isotone hypothesis (A)} \\
\implies \phi(x) \subset \phi(y) & \text{ by linear hypothesis (B)} \\
\end{align*}
\]

Lemma 1.4. Let $X \triangleq (X, \vee, \wedge; \leq)$ and $Y \triangleq (Y, \sqcup, \sqcap; \sqsubseteq)$ be lattices \(^{42}\) (Definition 1.33 page 12). Let $f$ be a function in $X^Y$. Let $\phi$ be a function in $Y^X$ \(^{42}\) (Definition 1.6 page 6).

\[
\begin{align*}
\{ \begin{array}{l}
A. \ \phi \text{ is strictly isotone} \quad (\text{Definition 1.36 page 16}) \text{ and } \\
B. \ X \text{ is linearly ordered} \quad (\text{Definition 1.25 page 10}) \text{ and } \\
C. \ Y \text{ is linearly ordered} \quad (\text{Definition 1.25 page 10})
\end{array} & \} \\
\implies \{ 1. \ \bigcup_{x \in X} \phi[f(x)] = \phi \bigg[ \bigvee_{x \in X} f(x) \bigg] \quad \text{and } \\
2. \ \bigcap_{x \in X} \phi[f(x)] = \phi \bigg[ \bigwedge_{x \in X} f(x) \bigg] \}
\end{align*}
\]

\[\text{42} \ \text{Burris and Sankappanavar (2000), page 10}\]
PROOF:

\[
\phi \left( \bigvee_{x \in X} f(x) \right) = \phi[\{f(a) | f(x) \leq f(a) \ \forall x, a \in X \}] \quad \text{because } f \in X^X \text{ and by Lemma 1.1 page 11}
\]

\[
= \{\phi[f(a)] | f(x) \leq f(a) \ \forall x \in X \}
\]

\[
= \{\phi[f(a)] | \phi[f(x)] \subseteq \phi[f(a)] \ \forall x \in X \} \quad \text{by Lemma 1.3 (page 17)}
\]

\[
\phi \left( \bigvee_{x \in X} f(x) \right) = \phi[f(a)] \quad \text{by definition of } (Y, \cup, \sqsubset; \subseteq)
\]

\[
\phi \left( \bigwedge_{x \in X} f(x) \right) = \phi[\{f(a) | f(x) \leq f(a) \ \forall x \in X \}] \quad \text{because } f \in X^X \text{ and by Lemma 1.1 page 11}
\]

\[
= \{\phi[f(a)] | f(x) \leq f(a) \ \forall x \in X \}
\]

\[
= \{\phi[f(a)] | \phi[f(x)] \subseteq \phi[f(a)] \ \forall x \in X \} \quad \text{by Lemma 1.3 (page 17)}
\]

\[
\phi \left( \bigwedge_{x \in X} f(x) \right) = \phi[f(x)] \quad \text{by definition of } (Y, \cup, \sqsubset; \subseteq)
\]

Lemma 1.5. Let \(X \triangleq (X, \vee, \wedge; \leq)\) and \(Y \triangleq (Y, \cup, \sqsubset; \subseteq)\) be LATTICES. Let \(f\) be a function in \(X^X\). Let \(\psi\) be a function in \(Y^X\).

A. \(\psi\) is STRICTLY ANTITONE (Definition 1.36 page 16) and
B. \(X\) is LINEARLY ORDERED (Definition 1.25 page 10) and
C. \(Y\) is LINEARLY ORDERED (Definition 1.25 page 10)

\[
\begin{align*}
\{ & \psi \left( \bigvee_{x \in X} f(x) \right) = \psi[f(x)] \quad \text{by definition of } (X, \vee, \wedge; \leq) \\
& \bigvee_{x \in X} \psi[f(x)] \quad \text{by definition of strictly antitone (Definition 1.36 page 16)} \\
& \bigwedge_{x \in X} \psi[f(x)] \quad \text{by definition of } (Y, \cup, \sqsubset; \subseteq) \\
\} \quad \Rightarrow \quad \left\{ \begin{array}{l}
1. \quad \bigvee_{x \in X} \psi[f(x)] \\
2. \quad \bigwedge_{x \in X} \psi[f(x)]
\end{array} \right. \quad \text{and}
\]

\[
\begin{align*}
\{ & \psi \left( \bigwedge_{x \in X} f(x) \right) = \psi[f(x)] \quad \text{by definition of } (X, \vee, \wedge; \leq) \\
& \bigwedge_{x \in X} \psi[f(x)] \quad \text{by definition of strictly antitone (Definition 1.36 page 16)} \\
& \bigvee_{x \in X} \psi[f(x)] \quad \text{by definition of } (Y, \cup, \sqsubset; \subseteq) \\
\} \quad \Rightarrow \quad \left\{ \begin{array}{l}
1. \quad \bigwedge_{x \in X} f(x) \\
2. \quad \bigvee_{x \in X} f(x)
\end{array} \right. \quad \text{and}
\]

Lemma 1.6. Let \(X \triangleq (X, \vee, \wedge; \leq)\) and \(Y \triangleq (Y, \cup, \sqsubset; \subseteq)\) be LATTICES (Definition 1.33 page 12). Let \(f\) be a function in \(X^X\). Let \(\phi\) and \(\psi\) be functions in \(Y^X\).

\[
\begin{align*}
\{ & \phi \text{ is STRICTLY ISOTONE (Definition 1.36 page 16) and } \\
& X \text{ is LINEARLY ORDERED (Definition 1.25 page 10) } \\
\} \quad \Rightarrow \quad \left\{ \begin{array}{l}
1. \quad \arg \bigvee_{x \in X} f(x) = \arg \bigvee_{x \in X} \phi[f(x)] \quad \text{and} \\
2. \quad \arg \bigwedge_{x \in X} f(x) = \arg \bigwedge_{x \in X} \phi[f(x)]
\end{array} \right.
\]

VERSION 0.50E

A thesis concerning symbolic sequence processing

Daniel J. Greenhoe 2016 Jun 16 6:40AM UTC
Proof:
\[
\begin{align*}
\arg \bigvee_{x \in X} f(x) &\triangleq \arg \bigvee \{f(x) \mid x \in X\} \\
&= \arg \{f(a) \mid f(x) \leq f(a) \quad \forall x, a \in X\} \\
&= \arg \{f(a) \mid \phi[f(a)] \leq \phi[f(a)] \quad \forall x, a \in X\} \\
&= \arg \{\phi[f(a)] \mid f(x) \leq f(a) \quad \forall x, a \in X\} \\
&= \bigvee_{x \in X} \phi[f(x)]
\end{align*}
\]
because \( f \in X^X \) and by Lemma 1.1 page 11
by hypothesis (A) and Lemma 1.5 page 18
because \( \arg \{f(x) \mid P(x)\} = \arg \{g[f(x)] \mid P(x)\} \)
because \( f \in X^X \) and by Lemma 1.1 page 11

Proof:
\[
\begin{align*}
\arg \bigwedge_{x \in X} f(x) &\triangleq \arg \bigwedge \{f(x) \mid x \in X\} \\
&= \arg \{f(a) \mid f(x) \leq f(a) \quad \forall x, a \in X\} \\
&= \arg \{f(a) \mid \phi[f(a)] \leq \phi[f(a)] \quad \forall x, a \in X\} \\
&= \arg \{\phi[f(a)] \mid f(x) \leq f(a) \quad \forall x, a \in X\} \\
&= \bigwedge_{x \in X} \phi[f(x)]
\end{align*}
\]
because \( f \in X^X \) and by Lemma 1.1 page 11
by hypothesis (A) and Lemma 1.5 page 18
because \( \arg \{f(x) \mid P(x)\} = \arg \{g[f(x)] \mid P(x)\} \)
because \( f \in X^X \) and by Lemma 1.1 page 11

Remark 1.2. Using the definitions of Lemma 1.6 (page 18), and letting \( g \) be a function in \( X^X \), and despite the results of Lemma 1.6, note that
A. \( \phi \) is strictly isotone
B. \( X \) is linearly ordered
For example, let \( f(x) \triangleq x, g(x) \triangleq -x + 2 \), and \( \phi(x) \triangleq x^2 \). Then
\[
\begin{align*}
\arg \bigvee_{x \in X} \{f(x)g(x)\} &\triangleq \arg \bigwedge \{-x^2 + 2x\} = 1 = \frac{4}{3} = \arg \bigwedge \{-x^3 + 2x^2\} = \arg \bigwedge_{x \in X} \phi[f(x)]g(x)
\end{align*}
\]

Lemma 1.7. Let \( X \triangleq (\mathcal{X}, \lor, \land, \leq) \) and \( Y \triangleq (\mathcal{Y}, \lor, \land, \subseteq) \) be LATTICES (Definition 1.33 page 12). Let \( f \) be a function in \((X \times X)^X\). Let \( \phi \) and \( \psi \) be functions in \( Y^X \).
A. \( \phi \) is strictly isotone
B. \( X \) is linearly ordered
\[
\begin{align*}
\text{arg} \bigwedge_{x \in X} \{f(x, y)\} &= \text{arg} \bigvee_{x \in X} \phi[f(x) \cup y] \quad \forall x, y \in X \quad \text{and} \\
\text{arg} \bigvee_{x \in X} \{f(x, y)\} &= \text{arg} \bigwedge_{x \in X} \phi[f(x) \cap y] \quad \forall x, y \in X \quad \text{and} \\
\text{arg} \bigwedge_{x \in X} \{f(x, y)\} &= \text{arg} \bigvee_{x \in X} \phi[f(x) \cup y] \quad \forall x, y \in X \quad \text{and}
\end{align*}
\]

Proof:
\[
\begin{align*}
\text{arg} \bigwedge_{x \in X} \{f(x, y)\} &= \text{arg} \bigwedge_{x \in X} \phi \left( \bigvee_{y \in X} f(x, y) \right) \quad \text{by Lemma 1.4 (page 17)} \\
&= \text{arg} \bigvee_{x \in X} f(x, y) \quad \text{by Lemma 1.2 (page 17)}
\end{align*}
\]
\[
\begin{align*}
\text{arg} \bigvee_{x \in X} \{f(x, y)\} &= \text{arg} \bigvee_{x \in X} \phi \left( \bigwedge_{y \in X} f(x, y) \right) \quad \text{by Lemma 1.4 (page 17)} \\
&= \text{arg} \bigwedge_{x \in X} f(x, y) \quad \text{by Lemma 1.2 (page 17)}
\end{align*}
\]
\[
\begin{align*}
\text{arg} \bigwedge_{x \in X} \{f(x, y)\} &= \text{arg} \bigwedge_{x \in X} \phi \left( \bigvee_{y \in X} f(x, y) \right) \quad \text{by Lemma 1.4 (page 17)} \\
&= \text{arg} \bigvee_{x \in X} f(x, y) \quad \text{by Lemma 1.2 (page 17)}
\end{align*}
\]
1.6. METRIC SPACE CONCEPTS

1.6.1 Motivation

Why might we care about metrics, or more generally distance functions, in symbolic sequence processing? Metric balls in a metric space induce a topology. Topologies are necessary for the concept of convergence. Some topologies are also algebra of sets; an algebra of sets (or in particular a sigma-algebra) is used for the concept of measure. Loosely speaking then, we care about distance and distance spaces for two reasons:

1. In analysis, metric spaces allow us to define the concepts of convergence and limit of a sequence as in \( \sum_{n=0}^{\infty} x_n \triangleq \lim_{N \to \infty} \sum_{n=0}^{N} x_n \). That is, without the implicit or explicit definition of convergence and limit, the expression \( \sum_{n=0}^{\infty} x_n \) is meaningless.

2. In signal processing, “optimal” decisions may be made with respect to a distance space. For example, a point may be selected (identified as “optimal”) based on it being measured as having the smallest distance to some reference point.

1.6.2 Isometric spaces

**Definition 1.37.** Let \((X, d)\) and \((Y, p)\) be distance spaces (Definition B.1 page 125). The function \(f \in Y^X\) is an isometry on \((Y, p)^{(X, d)}\) if 
\(d(x, y) = p(f(x), f(y))\) \(\forall x, y \in X\). The spaces \((X, d)\) and \((Y, p)\) are isometric if there exists an isometry on \((Y, p)^{(X, d)}\).

\(^{43}\)For example on the three element set \(\{x, y, z\}\), there are a total of 29 topologies; on of these 29, five are algebras of sets. References: \(\checkmark\) Isham (1999), page 44, \(\checkmark\) Isham (1989), page 1516, \(\checkmark\) Steiner (1966), page 386, \(\checkmark\) Sloane (2014) (http://oeis.org/A000798), \(\checkmark\) Brown and Watson (1996), page 31, \(\checkmark\) Comtet (1974) page 229, \(\checkmark\) Comtet (1966), \(\checkmark\) Chatterji (1967), page 7, \(\checkmark\) Evans et al. (1967), \(\checkmark\) Krishnamurthy (1966), page 157

\(^{44}\)Klauder (2010) page 4


\(\checkmark\)
Theorem 1.6. \(^{46}\) Let \((X, d)\) and \((Y, p)\) be distance spaces. Let \(f\) be a function in \(Y^X\) and \(f^{-1}\) its inverse in \(X^Y\).

\[
\{ f \text{ is an isometry on } (Y, p)(X, d) \} \iff \{ f^{-1} \text{ is an isometry on } (X, d)(Y, p) \}
\]

If a function \(p\) is a metric and a function \(g\) is injective, then the function \(d(x, y) \triangleq p(g(x), g(y))\) is also a metric (next theorem). For an example of this with \(p(x, y) \triangleq |x - y|\) and \(g \triangleq \arctan(x)\), see Example 1.7 (page 21).

Theorem 1.7 (Pullback metric/g-transform metric). \(^{47}\) Let \(X\) and \(Y\) be sets.

\[
\begin{align*}
(1). & \quad p \text{ is a metric on } Y \\
(2). & \quad g \text{ is a function in } Y^X \\
(3). & \quad g \text{ is injective}
\end{align*}
\]

\[
\implies \forall x, y \in X, \quad d(x, y) = p(g(x), g(y))
\]

\(d\) is a metric on \(X\)

\(\Box\) Proof:

1. Proof that \(x = y \implies d(x, y) = 0:\)

\[
d(x, y) \triangleq p(\phi(x), \phi(y)) = 0
\]

by definition of \(d\)

2. Proof that \(x = y \iff d(x, y) = 0:\)

\[
0 = d(x, y) \iff p(\phi(x), \phi(y)) = 0
\]

by right hypothesis

3. Proof that \(d(x, y) \leq d(z, x) + d(z, y):\)

\[
d(x, y) \leq \left( p(\phi(x), \phi(z)) + d(\phi(z), \phi(y)) \right)
\]

by subadditive property of \(p\)

\[
d(x, y) \triangleq d(z, x) + d(z, y)
\]

by definition of \(d\)

Example 1.7 (Inverse tangent metric). \(^{48}\) Let \((x_1, x_2, \ldots, x_N)\) and \((y_1, y_2, \ldots, y_N)\) be points in \(\mathbb{R}^N\).

\[
ds(\{x_1, x_2, \ldots, x_N\}, \{y_1, y_2, \ldots, y_N\}) \triangleq \sum_{n=1}^{N} |\arctan x_n - \arctan y_n|
\]

is a metric.

\(\Box\) Proof:

1. The function \(d(x, y) \triangleq |x - y|\) is a metric (the usual metric, Definition D.9 page 156).
2. The function \(g(x) \triangleq \arctan(x)\) is injective in \(\mathbb{R}\).
3. Therefore, \(d\) is a Pullback metric (or \(g\)-transform metric), and by Theorem 1.7 (page 21), \(d\) is a metric.

\(^{46}\) Thron (1966), page 153 (theorem 19.5)

\(^{47}\) Deza and Deza (2009) page 81

\(^{48}\) Copson (1968), page 25, Khamisi and Kirk (2001) page 14
1.7 Ordered metric spaces

1.7.1 Definitions

Definition 1.38. A triple \( G \triangleq (X, \leq, d) \) is an ordered quasi-metric space if \((X, d)\) is a quasi-metric space (Definition D.6 page 153) and \( (X, \leq) \) is an ordered set (Definition 1.24 page 10).

- \( G \) is an ordered metric space if \( d \) is a metric (Definition D.7 page 153).
- \( G \) is an unordered quasi-metric space if \( \leq \) is a metric and \( \leq \) is the standard linear order relation (Definition 1.25 page 10) on \( \mathbb{R} \).

Remark 1.4. Note that the four structures defined in Definition 1.38 are not mutually exclusive. For example, by Definition 1.38,

\[
\{ \text{ unordered metric space}\} \subseteq \{ \text{ unordered quasi-metric space}\} \subseteq \{ \text{ ordered quasi-metric space}\}
\]

\[
\{ \text{ unordered metric space}\} \subseteq \{ \text{ ordered metric space}\} \subseteq \{ \text{ ordered quasi-metric space}\}
\]

Remark 1.5. The use of the quasi-metric rather than exclusive use of the more restrictive metric in Definition 1.38 is motivated by state machines, where metrics measuring distances between states are in some cases by nature non-symmetric. One such example is the linear congruential pseudorandom number generator (Example 2.19 page 62).

Remark 1.6. This text makes extensive reference to the real line (next definition). There are several ways to define the real line. In particular, there are many possible ordering relations on \( \mathbb{R} \) and several possible topologies on \( \mathbb{R} \).\(^{49}\) In fact, order and topology are closely related in that an order relation \( \leq \) (Definition 1.24 page 10) on a set always induces a topology (called the order topology / interval topology\(^{50}\)); and in the case of the real line, a topology induces an order structure up to the order relation’s dual (Definition 1.26 page 10).\(^{51}\) This text uses a fairly standard structure, as defined next.

Definition 1.39. The triple \((\mathbb{R}, |\cdot|, \leq)\) is the real line if \( \mathbb{R} \) is the set of real numbers (Definition 1.2 page 5), \( d(x, y) \triangleq |x - y| \) is the usual metric on \( \mathbb{R} \) (Definition D.9 page 156), and \( \leq \) is the standard linear order relation (Definition 1.25 page 10) on \( \mathbb{R} \).

---


\(^{51}\) Hocking and Young (1961) page 52 (2–5 The interval and the circle), Salzmann et al. (2007) pages 69–70 (5.75 Note: Ordering and topology on \( \mathbb{R} \), see also 5.10 Theorem page 36)
**Definition 1.40.** The triple \((\mathbb{Z}, \leq, |\cdot|)\) is the integer line if \(\mathbb{Z}\) is the set of integers (Definition 1.2 page 5), \(d(m, n) \triangleq |m – n|\) is the usual metric (Definition D.9 page 156) on \(\mathbb{R}\) restricted to \(\mathbb{Z}\), and \(\leq\) is the standard linear order relation on \(\mathbb{Z}\) as induced by Peano’s Axioms.\(^{52}\)

### 1.7.2 Examples

**Example 1.8.** The integer line (Definition 1.40 page 23) is an ordered metric space (Definition 1.38 page 22), and is illustrated in Figure 1.4 page 22 (A).

**Example 1.9.** The real line (Definition 1.39 page 22) is an ordered metric space (Definition 1.38 page 22), and is illustrated in Figure 1.4 page 22 (B).

**Example 1.10.** The complex plane \((\mathbb{C}, |\cdot|, \leq)\) is an ordered metric space (Definition 1.38 page 22) where \(\mathbb{C} \triangleq \mathbb{R}^2\) is the set of complex numbers, \(d(x, y) \triangleq |x – y| \triangleq \sqrt{\mathbb{R}x – \mathbb{R}y}^2 + (\mathbb{R}x – \mathbb{R}y)^2\), \(\mathbb{R}x \triangleq \mathbb{R} (a, b) \triangleq a \forall (a, b) \in \mathbb{C}\) (\(\mathbb{R}x\) is the real part of \(x\)), \(\mathbb{I}x \triangleq \mathbb{I} (a, b) \triangleq b \forall (a, b) \in \mathbb{C}\) (\(\mathbb{I}x\) is the imaginary part of \(x\)), and \(\leq\) is any order relation defined on \(\mathbb{C}\). Possible order relations include the coordinatewise order relation (Example 1.1 page 10), the lexicographical order relation (Example 1.2 page 10), and \(\leq\) \(\leq_{\varnothing}\) (in which case the complex plane is unordered). The complex plane is illustrated in Figure 1.4 page 22 (C).

**Example 1.11.** A 6 element ring \(\{(0, 1, 2, 3, 4, 5), d, \varnothing\}\) is an unordered metric space (Definition 1.38 page 22) where the metric is defined on a ring as illustrated in Figure 1.4 page 22 (D), with each line segment representing a distance of 1.

**Example 1.12.** A 6 element discrete metric \(\{(0, 1, 2, 3, 4, 5), d, \varnothing\}\) is an unordered metric space (Definition 1.38 page 22) where the metric is the discrete metric (Definition D.8 page 155). This structure is illustrated in Figure 1.4 page 22 (E).

**Example 1.13.** Figure 1.4 page 22 (F) illustrates a linear congruential pseudo-random number generator induced by the equation \(y_{n+1} = (y_n + 2) \bmod 5 \text{ with } y_0 = 1\). The structure is an unordered quasi-metric space. See Example 2.17 (page 59)–Example 2.19 (page 62) for further demonstration.

### 1.8 Sequences

#### 1.8.1 Sequences

**Definition 1.41.** \(^{53}\) A function in \(X^D\) (Definition 1.6 page 6) is an \(X\)-valued sequence if \(D \neq \varnothing\) and \(D\) is a convex (Definition 1.29 page 11) subset of \(\mathbb{Z}\). A sequence may be denoted in the form \(\langle x_n \rangle_{n \in D}\), or simply as \(\langle x_n \rangle\).

**Definition 1.42.** The sequence \(\langle y_n \rangle_{D_2}\) is the sequence \(\langle x_n \rangle_{D_1}\) down sampled by a factor of \(M\), where \(M \in \mathbb{N}\), if \(n \in D_2 \iff Mn \in D_1\) and \(y_n = x_{Mn} \forall n \in D_2\).

**Definition 1.43.** Let \(\oplus\) be the addition operator (Definition 1.15 page 7) and \(\otimes\) the multiplication operator (Definition 1.16 page 7). Let \(D_1\) and \(D_2\) be convex subsets of \(\mathbb{Z}\).

Let \(D \triangleq (\bigwedge D_1 + \bigwedge D_2 - 1 : \bigvee D_1 + \bigvee D_2 + 1)\).

Let \(\langle x_n \rangle_{D_1}\) be a sequence over a field \(\mathbb{F}_1\) and \(\langle y_n \rangle_{D_2}\) a sequence over a field \(\mathbb{F}_2\).

The convolution \(\langle z_n \rangle_{D} \triangleq \langle x_n \rangle_{D_1} \ast \langle y_n \rangle_{D_2}\) of \(\langle x_n \rangle\) and \(\langle y_n \rangle\) is defined as

\[
z_n \triangleq \bigoplus_{m \in D_1} f(n, m) \text{ where } f \text{ is defined as } f(n, m) \triangleq \begin{cases} x_m \otimes y_{n-m} & \text{if } m \in D_1 \text{ and } (n-m) \in D_2 \\ 0 & \text{otherwise} \end{cases} \forall n, m \in D\]

---


\(^{53}\) Simmons (2016) (‘‘Formal definition’’)
Proposition 1.4. Let \((x_n)\) and \((y_n)\) be finite sequences with lengths \(N\) and \(M\), respectively. Then the length of \((x_n) \ast (y_n)\) is \(N + M - 1\).

Example 1.14. \(^{54}\) Let \((x_n)_{n \in \mathbb{Z}}\) and \((y_n)_{n \in \mathbb{Z}}\) be sequences over a field \(\mathbb{F}\). Then the domain \(\mathcal{D}\) of the convolution \((z_n)_{n \in \mathbb{D}} \triangleq (x_n) \ast (y_n)\) is
\[
\mathcal{D} \triangleq \left( \bigwedge Z + \bigwedge Z - 1 : \bigvee Z + \bigvee Z + 1 \right) = Z \text{ and } z_n \triangleq \sum_{m \in \mathbb{Z}} x_m y_{n-m}.
\]

Example 1.15. Let \((x_n)_{[0:1]} \triangleq (1, 2)\) and \((y_n)_{[0:2]} \triangleq (10, 20, 50)\) be sequences over the field \((\mathbb{R}, +, \cdot)\). Then the domain \(\mathcal{D}\) of the convolution \((z_n)_{n \in \mathbb{D}} \triangleq (x_n) \ast (y_n)\) is
\[
\mathcal{D} \triangleq (0 + 0 - 1 : 1 + 2 + 1) = [0 : 3] = \{0, 1, 2, 3\} \text{ and }
\]
\[
(z_n)_{n \in \mathcal{D}} \triangleq \left( \sum_{m \in \{0,1\}} f(0, m), \sum_{m \in \{0,1\}} f(1, m), \sum_{m \in \{0,1\}} f(2, m), \sum_{m \in \{0,1\}} f(3, m) \right)_{[0,1,2,3]}
\]
\[
\triangleq \left( (1 \times 10 + 0), (1 \times 20 + 2 \times 10), (1 \times 50 + 2 \times 20), (0 + 2 \times 50) \right)_{[0,1,2,3]}
\]
\[
= \left( \frac{10}{z_0}, \frac{40}{z_1}, \frac{90}{z_2}, \frac{100}{z_3} \right)_{[0,1,2,3]}
\]

Example 1.16. Let \((x_n)_{[0:1]} \triangleq (1, 2)\) and \((y_n)_{[3:5]} \triangleq (10, 20, 50)\) be sequences over the field \((\mathbb{R}, +, \cdot)\). Then the domain \(\mathcal{D}\) of the convolution \((z_n)_{n \in \mathcal{D}} \triangleq (x_n) \ast (y_n)\) is
\[
\mathcal{D} \triangleq (0 + 3 - 1 : 1 + 5 + 1) = [3 : 6] = \{3, 4, 5, 6\} \text{ and }
\]
\[
(z_n)_{n \in \mathcal{D}} \triangleq \left( \sum_{m \in \{0,1\}} f(3, m), \sum_{m \in \{0,1\}} f(4, m), \sum_{m \in \{0,1\}} f(5, m), \sum_{m \in \{0,1\}} f(6, m) \right)_{[3,4,5,6]}
\]
\[
\triangleq \left( (1 \times 10 + 0), (1 \times 20 + 2 \times 10), (1 \times 50 + 2 \times 20), (0 + 2 \times 50) \right)_{[3,4,5,6]}
\]
\[
= \left( \frac{10}{z_3}, \frac{40}{z_4}, \frac{90}{z_5}, \frac{100}{z_6} \right)_{[3,4,5,6]}
\]

Example 1.17. Let \((x_n)_{[0:1]} \triangleq (1, 2)\) and \((y_n)_{[0:2]} \triangleq (3, 4, 5, 6, 7, 8)\) be sequences. Then the domain \(\mathcal{D}\) of the convolution \((z_n)_{n \in \mathcal{D}} \triangleq (x_n) \ast (y_n)\) is \(\mathcal{D} = \{0, 1, 2, 3\}\) and
\[
(z_n)_{n \in \mathcal{D}} \triangleq \left( \bigoplus_{m \in \{0,1\}} f(0, m), \bigoplus_{m \in \{0,1\}} f(1, m), \bigoplus_{m \in \{0,1\}} f(2, m), \bigoplus_{m \in \{0,1\}} f(3, m) \right)_{[0,1,2,3]}
\]
\[
\triangleq \left( \{1 \otimes (3, 4) \oplus 0\}, \{1 \otimes (5, 6) \oplus 2 \otimes (3, 4)\}, \{1 \otimes (7, 8) \oplus 2 \otimes (5, 6)\}, \{0 \oplus 2 \otimes (7, 8)\} \right)_{[0,1,2,3]}
\]
\[
\triangleq \left( \{(3, 4), (5, 6) \oplus (6, 8), (7, 8) \oplus (10, 12), (14, 16)\} \right)_{[0,1,2,3]}
\]
\[
= \left( \frac{3, 4}{z_0}, \frac{11, 14}{z_1}, \frac{17, 20}{z_2}, \frac{14, 16}{z_3} \right)_{[0,1,2,3]}
\]

Definition 1.44. Let \((x_n)_{n \in \mathbb{F}}\) and \((y_n)_{n \in \mathbb{F}}\) be sequences over a field \(\mathbb{F} \triangleq (X, +, \cdot)\), and \(a\) an element in \(\mathbb{F}\). The operations \(a + \langle x_n \rangle\), \((x_n) + a\), \(a \langle x_n \rangle\), and \(\langle x_n \rangle a\) are defined as
\[
\alpha \triangleq \langle x_n \rangle + a \triangleq \langle x_n + a \rangle_{n \in \mathbb{F}} \text{ and } \alpha \triangleq \langle x_n \rangle a_{n \in \mathbb{F}}
\]

\(^{54}\)Historical references: \(\equiv\) Cauchy (1821) (Chapter IV), \(\equiv\) Apostol (1975) page 204 (note that convolution is a single element in a series that is the “Cauchy product”), \(\equiv\) Dominguez-Torres (2010) page 20 (section 4.2: connection to the work of Cauchy), \(\equiv\) Dominguez-Torres (2015) (history of the continuous convolution operation)
1.8.2 Filtering

Definition 1.45. Let \( \{x_n\}_{n \in \mathbb{D}} \) and \( \{y_n\}_{n \in \mathbb{D}} \) be sequences (Definition 1.41 page 23). The sequence \( \{z_n\}_{n \in \mathbb{D}} \) is said to be \( \{y_n\}_n \) filtered by \( \{y_n\}_n \) if \( [z_n] \triangleq \{x_n\} \ast \{y_n\} \) (Definition 1.43 page 23). Moreover, in this case, the operation \( \ast \{y_n\}_n \) is a filter on the sequence \( \{x_n\}_n \).

Definition 1.46. A length \( M \) low pass rectangular sequence \( \{h_n\}_{n \in \{0, M-1\}} \) is here defined as
\[
h_n = \begin{cases} 
0 & \text{for } n = 0 \\
(-1)^{n+1} & \text{for } n = 1, 2, \ldots, M 
\end{cases}
\]

Note that in this definition, the sequence has been offset by 1 on the x-axis from what might normally be expected. This is for the purpose of computational convenience used in Section 3.3.2 (page 91).

Example 1.18. A length 16 low pass rectangular sequence and length 16 high pass rectangular sequence are illustrated below:

Definition 1.47. A length \( M \) high pass rectangular sequence \( \{h_n\}_{n \in \{0, M-1\}} \) is here defined as
\[
h_n = \begin{cases} 
0 & \text{for } n = 0 \\
(-1)^{n} & \text{for } n = 1, 2, \ldots, M 
\end{cases}
\]

Example 1.19. A length 50 low pass rectangular sequence is illustrated below:

Definition 1.48. A length \( M \) low pass Hanning sequence \( \{h_n\}_{n \in \{0, M-1\}} \) is here defined as
\[
h_n = \frac{1}{2} \left[ 1 - \cos \left( \frac{2\pi n}{M-1} \right) \right] \text{ for } n = 0, 1, 2, \ldots, M - 1.
\]

Example 1.19. A length 50 low pass Hanning sequence is illustrated below:

Definition 1.49. A length \( M \) high pass Hanning sequence \( \{h_n\}_{n \in \{0, M-1\}} \) is here defined as
\[
h_n = \frac{1}{2} \left[ 1 - \cos \left( \frac{2\pi n}{M-1} \right) \right] \text{ for } n = 0, 1, 2, \ldots, M - 1.
\]

Example 1.20. A length 50 high pass Hanning sequence is illustrated below:

Definition 1.50. A length \( M \) Haar scaling sequence \( \{h_n\}_{n \in \{0, M-1\}} \) is here defined as
\[
h_n = \sqrt{\frac{1}{M}} \text{ for } n \in \{0, M-1\}.
\]

Definition 1.51. A length \( M \) Haar wavelet sequence \( \{h_n\}_{n \in \{0, M-1\}} \) is here defined as
\[
h_n = \begin{cases} 
\pm \sqrt{\frac{1}{M}} & \text{ for } n = 0, 1, \ldots, \left\lfloor M/2 \right\rfloor - 1 \\
-\sqrt{\frac{1}{M}} & \text{ for } n = \left\lfloor M/2 \right\rfloor, \left\lfloor M/2 \right\rfloor + 1, \ldots, M - 1 \\
0 & \text{ otherwise}
\end{cases}
\]

---

Example 1.21. A length 8 Haar scaling sequence, a length 8 Haar wavelet sequence, and length 9 Haar wavelet sequence are illustrated below:

1.8.3 Discrete Fourier Transform

Definition 1.52. Let $\oplus$ be the addition operator (Definition 1.15 page 7) and $\otimes$ the multiplication operator (Definition 1.16 page 7). Let $(x_n)_{n \in \mathbb{D}}$ be a length $N$ sequence (Definition 1.41 page 23). The discrete Fourier transform $\mathcal{DFT}(x_n)$ of $(x_n)$ is a sequence $(y_k)_{k \in \mathbb{D}}$ over $\mathbb{C}$, where the element $y_k$ is defined as

$$y_k \triangleq \sqrt{\frac{1}{N}} \bigoplus_{n \in \mathbb{D}} x_n \otimes \exp \left( -\frac{i2\pi nk}{N} \right)$$
CHAPTER 2

STOCHASTIC PROCESSING ON WEIGHTED GRAPHS

“Les mathématiciens n’étudient pas des objets, mais des relations entre les objets; il leur est donc indifférent de remplacer ces objets par d’autres, pourvu que les relations ne changent pas. La matière ne leur importe pas, la forme seule les intéresse.”

Jules Henri Poincaré (1854-1912), physicist and mathematician

“Mathematicians do not study objects, but the relations between objects; to them it is a matter of indifference if these objects are replaced by others, provided that the relations do not change. Matter does not engage their attention, they are interested in form alone.”

2.1 Outcome subspaces

2.1.1 Definitions

Traditional probability theory is performed in a probability space \((\Omega, E, P)\). This section extends\(^2\) the probability space structure to include what herein is called an outcome subspace (next definition).

**Definition 2.1.** An extended probability space is the tuple \((\Omega, \preceq, d, E, P)\) where \((\Omega, E, P)\) is a probability space (Definition 1.21 page 8) and \((\Omega, d, \preceq)\) is an ordered quasi-metric space (Definition 1.38 page 22). The 4-tuple \((\Omega, \preceq, d, P)\) is an outcome subspace of the extended probability space \((\Omega, \preceq, d, E, P)\).

**Definition 2.2.** Let \(G \triangleq (\Omega, \preceq, d, P)\) be an outcome subspace (Definition 2.1 page 27).

The \textit{n-th}-moment \(m_n(x, y)\) from \(x\) to \(y\) in \(G\) is defined as
\[
m(x, y) \triangleq [d(x, y)]^n P(y) \quad \forall x, y \in \Omega, n \in \mathbb{N}.
\]

The \textit{moment} \(m(x, y)\) from \(x\) to \(y\) in \(G\) is defined as
\[
m(x, y) \triangleq m_n(x, y) \quad \forall x, y \in \Omega.
\]

This paper introduces a quantity called the \textit{outcome center} of an outcome subspace (next definition) which is in essence the same as the \textit{center of a graph} (Definition 1.19 page 8).

\(^1\) quote: \(\text{Poincaré (1902a) (Chapter 2)}\)
translation: \(\text{Poincaré (1902b), page 20}\)
image: \(\text{http://www-groups.dcs.st-and.ac.uk/~history/PictDisplay/Poincare.html}\)

\(^2\) Feldman and Valdez-Flores (2010) page 4 (“The name “random variable” is actually a misnomer, since it is not random and not a variable...the random variable simply maps each point (outcome) in the sample space to a number on the real line...Technically, the space into which the random variable maps the sample space may be more general than the real line...”)
Definition 2.3. Let $G \triangleq (\Omega, \preceq, \hat{d}, \hat{P})$ be an outcome subspace (Definition 2.1 page 27).

\[
\hat{\gamma}(G) \triangleq \operatorname{arg \ min} \max_{x \in \Omega} \sum_{y \in \Omega} d(x, y) P(y)
\]

is the outcome center of $G$.

The following additional definitions are of interest due in part to Corollary D.2 (page 167) and the minimax inequality (Theorem 1.2 page 12). They are illustrated in several examples in this section. However, most of them are not used outside this section.

Definition 2.4. Let $G \triangleq (\Omega, \preceq, \hat{d}, \hat{P})$ be an outcome subspace (Definition 2.1 page 27).

\[
\hat{\gamma}_a(G) \triangleq \operatorname{arg \ min} \sum_{y \in \Omega} d(x, y) P(y)
\]

is the arithmetic center of $G$.

\[
\hat{\gamma}_g(G) \triangleq \operatorname{arg \ min} \prod_{y \in \Omega} [d(x, y) P(y)]
\]

is the geometric center of $G$.

\[
\hat{\gamma}_h(G) \triangleq \operatorname{arg \ min} \left( \sum_{y \in \Omega} \frac{1}{d(x, y)} P(y) \right)^{-1}
\]

is the harmonic center of $G$.

\[
\hat{\gamma}_m(G) \triangleq \operatorname{arg \ max} \min_{y \in \Omega} d(x, y) P(y)
\]

is the minimal center of $G$.

\[
\hat{\gamma}_m(G) \triangleq \operatorname{arg \ max} \min_{y \in \Omega} d(x, y) P(y)
\]

is the maxmin center of $G$.

In a manner similar to the traditional variance function (Definition 1.23 page 9), the outcome variance (next) is a kind of measure of the quality of the outcome center as a representative estimate of all the values of the outcome subspace. Said another way, it is in essence the expected error of the center measure.

Definition 2.5. Let $G \triangleq (\Omega, \preceq, \hat{d}, \hat{P})$ be an outcome subspace (Definition 2.1 page 27).

The outcome variance $\operatorname{Var}(G; \hat{\gamma}_x)$ of $G$ with respect to $\hat{\gamma}_x$ is

\[
\operatorname{Var}(G; \hat{\gamma}_x) \triangleq \sum_{x \in \Omega} d^2(\hat{\gamma}_x(G), x) P(x).
\]

where $\hat{\gamma}_x$ is any of the operators defined in Definition 2.3 or Definition 2.4 (page 28).

Moreover, $\operatorname{Var}(G) \triangleq \operatorname{Var}(G; \hat{\gamma})$, where $\hat{\gamma}$ is the outcome center (Definition 2.3 page 28).

Remark 2.1. The quantity $p(x, \Omega) \triangleq \sum_{y \in \Omega} d(x, y) P(y)$ in the arithmetic center $\hat{\gamma}_a(G)$ (Definition 2.4 page 28) is itself a metric (Definition D.7 page 153). Thus, $\hat{\gamma}_a(G)$ is the $x$ that produces the minimum of all the metrics with center $x$.

Proof: This follows directly from power mean metrics theorem with $r = 1$ (Theorem D.10 page 159).

2.1.2 Specific outcome subspaces

Definition 2.6. The structure $G \triangleq (\{\emptyset, \emptyset, \emptyset, \emptyset, \emptyset\}, \hat{d}, \preceq, \hat{P})$ is the weighted die outcome subspace if $G$ is an outcome subspace, $\preceq = \emptyset$ (unordered Definition 1.24 page 10), and $\hat{d}$ is the discrete metric (Definition D.8 page 155).

Definition 2.7. The structure $G \triangleq (\{\emptyset, \emptyset, \emptyset, \emptyset, \emptyset\}, \hat{d}, \preceq, \hat{P})$ is the fair die outcome subspace if $G$ is a weighted die outcome subspace (Definition 2.6), and

\[
\hat{P}(\emptyset) = P(\emptyset) = P(\emptyset) = P(\emptyset) = P(\emptyset) = P(\emptyset) = \frac{1}{5}.
\]
Figure 2.1: example outcome subspaces (Definition 2.1 page 27) illustrated by weighted graphs with shaded expected values.

**Definition 2.8.** The structure $G \triangleq (\{\square, \circ, \Diamond, \bigcirc, \blacklozenge, \blacklozenge\}, \hat{d}, \emptyset, \hat{P})$ is the weighted real die outcome subspace if $G$ is an outcome subspace, and metric $\hat{d}$ is defined as in the table to the right.

**Definition 2.9.** The structure $G \triangleq (\{\square, \circ, \Diamond, \bigcirc, \blacklozenge, \blacklozenge\}, \hat{d}, \emptyset, \hat{P})$ is the real die outcome subspace if $G$ is a weighted real die outcome subspace (Definition 2.8 page 29) with

$$\hat{P}(\square) = \hat{P}(\circ) = \hat{P}(\Diamond) = \hat{P}(\bigcirc) = \hat{P}(\blacklozenge) = \frac{1}{6}.$$

**Definition 2.10.** The structure $G \triangleq (\{\updownarrow, \bigcirc, \bigcirc, \bigcirc, \bigcirc\}, \hat{d}, \emptyset, \hat{P})$ is the spinner outcome subspace if $G$ is an outcome subspace,

$$\hat{P}(\updownarrow) = \hat{P}(\bigcirc) = \hat{P}(\bigcirc) = \hat{P}(\bigcirc) = \hat{P}(\bigcirc) = \frac{1}{6},$$

and metric $\hat{d}$ is defined as in the table to the right.

**Definition 2.11.** The structure $H \triangleq (\{\square, \bigcirc, \bigcirc, \bigcirc, \bigcirc\}, \hat{d}, \emptyset, \hat{P})$ is the DNA outcome subspace, or genome outcome subspace, if $H$ is an outcome subspace, and $\hat{d}$ is the discrete metric (Definition D.8 page 155).

**Definition 2.12.** The structure $H \triangleq (\{\square, \bigcirc, \bigcirc, \bigcirc, \bigcirc\}, \hat{d}, \hat{d}, \hat{P})$ is the DNA scaffold outcome subspace, or genome scaffold outcome subspace, if $H$ is an outcome subspace,

$$\hat{d}(\square, \bigcirc) = \hat{d}(\bigcirc, \bigcirc) = \hat{d}(\bigcirc, \bigcirc) = \hat{d}(\bigcirc, \bigcirc) = \hat{d}(\bigcirc, \bigcirc) = \frac{1}{6},$$

and $\hat{P}$ is a probability function, and metric $\hat{d}$ is defined as in the table to the right.

### 2.1.3 Example calculations

**Example 2.1.** Let $G \triangleq (\{\square, \circ, \Diamond, \bigcirc, \blacklozenge, \blacklozenge\}, \hat{d}, \hat{P})$ be the fair die outcome subspace (Definition 2.7 page 28). This structure is illustrated by the weighted graph (Definition 1.18 page 8) in Figure 2.1 (page 29).
(A), where each line segment represents a distance of 1. This structure has the following geometric values:

\[ \hat{C}(G) = \hat{C}_{\hat{G}}(G) = \hat{C}_{\hat{G}}(G) = \hat{C}_{\hat{G}}(G) = \hat{C}_{\hat{G}}(G) = \hat{C}_{\hat{G}}(G) = \{ \square, \Diamond, \Box, \Box, \Box, \Box, \Box \} \]

\[ \operatorname{Var}(G) = \operatorname{Var}(G; \hat{C}_{\hat{G}}) = \operatorname{Var}(G; \hat{C}_{\hat{G}}) = \operatorname{Var}(G; \hat{C}_{\hat{G}}) = \operatorname{Var}(G; \hat{C}_{\hat{G}}) = \operatorname{Var}(G; \hat{C}_{\hat{G}}) = 0 \]

\[ \text{P. PROOF:} \]

\[ \hat{C}(G) \triangleq \operatorname{arg min}_{x \in G} \max_{y \in G} d(x, y) \operatorname{P}(y) \quad \text{by definition of } \hat{C} \text{ (Definition 2.3 page 28)} \]

\[ = \operatorname{arg min}_{x \in G} \max_{y \in G} d(x, y) \frac{1}{6} \quad \text{by definition of } G \]

\[ = \operatorname{arg min}_{x \in G} \frac{1}{6} \{ 1, 1, 1, 1, 1, 1 \} \quad \text{by definition of } \text{discrete metric} \text{ (Definition D.8 page 155)} \]

\[ = \{ \square, \square, \Box, \Box, \Box, \Box, \Box \} \quad \text{by definition of } G \]

\[ \hat{C}_{\hat{G}}(G) \triangleq \operatorname{arg min}_{x \in G} \sum_{y \in \hat{G}(x)} d(x, y) \operatorname{P}(y) \quad \text{by definition of } \hat{C}_{\hat{G}} \text{ (Definition 2.4 page 28)} \]

\[ = \operatorname{arg min}_{x \in G} \sum_{y \in \hat{G}(x)} d(x, y) \frac{1}{6} \quad \text{by definition of } G \]

\[ = \operatorname{arg min}_{x \in G} \frac{1}{6} \{ 5, 5, 5, 5, 5, 5 \} \quad \text{by definition of } \text{discrete metric} \text{ (Definition D.8 page 155)} \]

\[ = \{ \square, \square, \Box, \Box, \Box, \Box, \Box \} \quad \text{by definition of } G \]

\[ \hat{C}_{\hat{G}}(G) \triangleq \operatorname{arg min}_{x \in G} \prod_{y \in \hat{G}(x)} d(x, y) \operatorname{P}(y) \quad \text{by definition of } \hat{C}_{\hat{G}} \text{ (Definition 2.4 page 28)} \]

\[ = \operatorname{arg min}_{x \in G} \prod_{y \in \hat{G}(x)} \frac{1}{6} \quad \text{by definition of } G \]

\[ = \operatorname{arg min}_{x \in G} \left\{ \frac{6}{5}, \frac{6}{5}, \frac{6}{5}, \frac{6}{5}, \frac{6}{5}, \frac{6}{5} \right\} \quad \text{by definition of } G \]

\[ = \{ \square, \square, \Box, \Box, \Box, \Box, \Box \} \quad \text{by definition of } G \]

\[ \hat{C}_{\hat{G}}(G) \triangleq \operatorname{arg min}_{x \in G} \min_{y \in \hat{G}(x)} d(x, y) \operatorname{P}(y) \quad \text{by definition of } \hat{C}_{\hat{G}} \text{ (Definition 2.4 page 28)} \]

\[ = \operatorname{arg min}_{x \in G} \min_{y \in \hat{G}(x)} \frac{1}{6} \quad \text{by definition of } G \]

\[ = \operatorname{arg min}_{x \in G} \frac{1}{6} \{ 1, 1, 1, 1, 1 \} \quad \text{by definition of } G \]

\[ \hat{C}_{\hat{G}}(G) \triangleq \operatorname{arg max}_{x \in G} \min_{y \in \hat{G}(x)} d(x, y) \operatorname{P}(y) \quad \text{by definition of } \hat{C}_{\hat{G}} \text{ (Definition 2.4 page 28)} \]

\[ = \operatorname{arg max}_{x \in G} \min_{y \in \hat{G}(x)} \frac{1}{6} \quad \text{by } \hat{C}_{\hat{G}}(G) \text{ result} \]

\[ = \operatorname{arg max}_{x \in G} \frac{1}{6} \{ 1, 1, 1, 1, 1 \} \quad \text{by } \hat{C}_{\hat{G}}(G) \text{ result} \]

\[ = \{ \square, \square, \Box, \Box, \Box, \Box, \Box \} \quad \text{by definition of } G \]

\[ \operatorname{Var}(G; \hat{C}_{\hat{G}}) = \operatorname{Var}(G; \hat{C}_{\hat{G}}) = \operatorname{Var}(G; \hat{C}_{\hat{G}}) = \operatorname{Var}(G; \hat{C}_{\hat{G}}) = \operatorname{Var}(G; \hat{C}_{\hat{G}}) = \operatorname{Var}(G; \hat{C}_{\hat{G}}) = 0 \]
\[ \sum_{x \in G} \left[ d(\hat{\mathcal{C}}(G), x) \right]^2 P(x) \]

by definition of \( \mathcal{V} \) (Definition 2.5 page 28)

\[ = \sum_{x \in G} \left( 0^2 \right) \frac{1}{6} \]

because \( \hat{\mathcal{C}}(G) = G \)

\[ = 0 \]

by field property of additive identity element 0

Example 2.2. Let \( G \triangleq \{ \Box, \Box, \Box, \Box, \Box, \Box \}, \emptyset, \leq, P \) be the real die outcome subspace (Definition 2.9 page 29). This structure is illustrated by the weighted graph (Definition 1.18 page 8) in Figure 2.1 (page 29) (B), where each line segment represents a distance of 1. The structure has the following geometric values:

\[ \hat{\mathcal{C}}(G) = \hat{\mathcal{C}}_1(G) = \hat{\mathcal{C}}_2(G) = \hat{\mathcal{C}}_3(G) = \hat{\mathcal{C}}_4(G) = \hat{\mathcal{C}}_5(G) = \emptyset \]

\[ \mathcal{V} \mathcal{A}(G) = \mathcal{V} \mathcal{A}(G; \hat{\mathcal{C}}_1) = \mathcal{V} \mathcal{A}(G; \hat{\mathcal{C}}_2) = \mathcal{V} \mathcal{A}(G; \hat{\mathcal{C}}_3) = \mathcal{V} \mathcal{A}(G; \hat{\mathcal{C}}_4) = \mathcal{V} \mathcal{A}(G; \hat{\mathcal{C}}_5) = 0 \]

\[ \emptyset \]

**Proof:**

\[ \hat{\mathcal{C}}(G) \triangleq \arg \min_{x \in G} \max_{y \in G} d(x, y) P(y) \]

by definition of \( \hat{\mathcal{C}} \) (Definition 2.3 page 28)

\[ = \arg \min_{x \in G} \max_{y \in G} d(x, y) \frac{1}{6} \]

by definition of \( G \)

\[ = \arg \min_{x \in G} \frac{1}{6} \{ 2, 2, 2, 2, 2, 2 \} \]

because for each \( x \), there is a \( y \) such that \( d(x, y) = 2 \)

\[ = \{ \Box, \Box, \Box, \Box, \Box, \Box \} \]

by definition of \( G \)

\[ \hat{\mathcal{C}}_1(G) \triangleq \arg \min_{x \in G} \sum_{y \in G} d(x, y) P(y) \]

by definition of \( \hat{\mathcal{C}}_1 \) (Definition 2.4 page 28)

\[ = \arg \min_{x \in G} \sum_{y \in G} d(x, y) \frac{1}{6} \]

by definition of \( G \)

\[ = \arg \min_{x \in G} \frac{1}{6} \left\{ \begin{array}{c} 0 + 1 + 1 + 1 + 1 + 2 \\
1 + 1 + 1 + 1 + 1 + 2 \\
1 + 1 + 0 + 1 + 2 + 1 \\
1 + 1 + 0 + 1 + 2 + 1 \\
1 + 1 + 1 + 1 + 1 + 0 + 1 \\
2 + 1 + 1 + 1 + 1 + 0 + 1 \\end{array} \right\} = \arg \min_{x \in G} \frac{1}{6} \left\{ \begin{array}{c} 6 \\
6 \\
6 \\
6 \\
6 \\
6 \end{array} \right\} = \{ \Box, \Box, \Box, \Box, \Box, \Box \} \]

\[ \hat{\mathcal{C}}_k(G) \triangleq \arg \min_{x \in G} \prod_{y \in G(x)} d(x, y)^{P(y)} \]

by definition of \( \hat{\mathcal{C}}_k \) (Definition 2.4 page 28)

\[ = \arg \min_{x \in G} \prod_{y \in G(x)} d(x, y)^{\frac{1}{6}} \]

by definition of \( G \)

\[ = \arg \min_{x \in G} \prod_{y \in G(x)} d(x, y)^{\frac{1}{6}} \]

because \( f(x) = x^{\frac{1}{6}} \) is strictly isotone and by Lemma 1.2 (page 17)

\[ = \arg \min_{x \in G} \prod_{y \in G(x)} d(x, y)^{\frac{1}{6}} \]

\[ = \{ 2, 2, 2, 2, 2, 2 \} \]

\[ = \{ \Box, \Box, \Box, \Box, \Box, \Box \} \]

\[ \hat{\mathcal{L}}(X) \triangleq \arg \min_{x \in G} \left( \sum_{y \in G(x)} \frac{1}{d(x, y)} P(y) \right)^{-1} \]

by definition of \( \hat{\mathcal{L}} \) (Definition 2.4 page 28)

\[ = \arg \min_{x \in G} \left( \sum_{y \in G(x)} \frac{1}{d(x, y)} \frac{1}{6} \right)^{-1} \]

by definition of \( G \)
\[
\hat{t}_m(G) \triangleq \arg \min_{x \in G} \min_{y \in \Omega(x)} d(x, y) P(y) \quad \text{by definition of } \hat{t}_m \text{ (Definition 2.4 page 28)}
\]

\[
= \arg \min_{x \in G} \min_{y \in \Omega(x)} \frac{1}{6} \quad \text{by definition of } \hat{t}_m \text{ (Definition 2.3 page 28)}
\]

\[
= \{ [2, 2, 2, 2, 2, 2] \}
\]

\[
\hat{t}_m(G) \triangleq \arg \max_{x \in G} \min_{y \in \Omega(x)} d(x, y) P(y) \quad \text{by definition of } \hat{t}_m \text{ (Definition 2.4 page 28)}
\]

\[
= \frac{1}{6} \{ [2, 2, 2, 2, 2, 2] \} \quad \text{by } \hat{t}_m(G) \text{ result}
\]

\[
\mathbb{V} \mathbb{R}(G; \hat{t}_m) = \mathbb{V} \mathbb{R}(G; \hat{t}_n) = \mathbb{V} \mathbb{R}(G; \hat{t}_m) = \mathbb{V} \mathbb{R}(G; \hat{t}_n) = \mathbb{V} \mathbb{R}(G)
\]

\[
\sum_{x \in G} [d(\hat{t}(G), x)]^2 P(x) \quad \text{by definition of } \mathbb{V} \mathbb{R} \text{ (Definition 2.5 page 28)}
\]

\[
\sum_{x \in G} (0) \frac{1}{6} \quad \text{because } \hat{t}(G) = G
\]

\[
= 0 \quad \text{by field property of additive identity element 0}
\]

\[\square\]

**Remark 2.2.** Let \( G \triangleq (\Omega, d, \leq, P) \) be the *fair die outcome subspace* (Example 2.1 page 29). Let \( H \triangleq (\Omega, p, \leq, P) \) be the *real die outcome subspace* (Example 2.2 page 31). These two subspaces are identical except for their metrics \( d \) and \( p \). So we can say that \( G \) and \( H \) are distinguished by their metrics. However, note that they are *indistinguishable* by the topologies induced by their metrics, because they both induce the same topology—the *discrete topology* \( 2^\Omega \) (Definition 1.8 page 6). That is, the geometric distinction provided in metric spaces is in general lost in topological spaces. Thus, topological spaces are arguably too general for the type of stochastic processing presented in this paper; rather, the stochastic processing discussed in this paper calls for metric space structure. And in this paper, this type of metric space structure is referred to as *metric geometry.*

\[\square\]

**Proof:**

1. Every metric space \((\Omega, d)\) (Definition 0.7 page 153) induces a *topological space* \((\Omega, T)\).

2. In particular, a metric \( d \) induces an *open ball* \( B(x, r) \in (2^\Omega)^{\omega \times R^+} \) centered at \( x \) with radius \( r \) such that \( B(x, r) \triangleq \{ y \in \Omega | d(x, y) < r \} \).

3. At each outcome \( x \) in \( G \), only two *open balls* are possible: \( B(x, r) = \{ \{ x \} \} \) for \( 0 < r \leq 1 \).

4. Let \( x' \) represent the die face which, when its numeric value is summed “in the usual way” with the numeric value of the die face \( x \), equals 7. Then at each point \( x \) in \( H \), three *open balls* are possible:

\[
B(x, r) = \begin{cases} 
\{ x \} & \text{for } 0 < r \leq 1 \\
\{ x, x' \} & \text{for } 1 < r \leq 2 \\
\{ \Omega \} & \text{for } r > 2 
\end{cases}
\]
5. The open balls of $(\Omega, d)$ or $(\Omega, p)$ in turn induce a base for a topology $T$, such that $T = \{ U \in 2^\Omega | U$ is a union of open balls $\}$. The topology induced by $G$ is the discrete topology $2^\Omega$ (Definition 1.8 page 6). The topology induced by $H$ is also the discrete topology $2^\Omega$.

6. So the metrics of $G$ and $H$ are different. And the balls induced by $G$ and those induced by $H$ are different. However, the topologies induced by $G$ and $H$ are the same.

Example 2.3. The weighted real die outcome subspace (Definition 2.6 page 28) illustrated in Figure 2.1 page 29 (C) has the following geometric characteristics:

$$
\hat{C}(G) = \hat{C}_b(G) = \{ \{ \} \} \quad \text{Var}(G) = \frac{101}{120} \approx 0.33
$$

$$
\hat{C}_\lambda(G) = \{ \{ \}, \{ \} \} \quad \text{Var}(G; \hat{C}_\lambda) = \frac{127}{150} \approx 0.847
$$

$$
\hat{C}_\mu(G) = \{ \{ \}, \{ \}, \{ \} \} \quad \text{Var}(G; \hat{C}_\mu) = \frac{145}{150} \approx 0.967
$$

$$
\hat{C}_m(G) = \{ \{ \}, \{ \}, \{ \}, \{ \} \} \quad \text{Var}(G; \hat{C}_m) = \frac{11}{30} = 0.22
$$

Note that the outcome center $\hat{C}(G)$ and arithmetic center $\hat{C}_b(G)$ again yield identical results. Also note that the four center measures of cardinality 1 ( $| \hat{C}(G) | = | \hat{C}_b(G) | = | \hat{C}_\lambda(G) | = | \hat{C}_m(G) | = 1$ Definition 1.13 page 6), $\hat{C}$ and $\hat{C}_b$ yield by far the lowest variance measures.

Proof:

$\hat{C}(G) \triangleq \arg \min_{x \in G} \max_{y \in G} d(x, y) P(y)$ by definition of $\hat{C}$ (Definition 2.3 page 28)

$$
\hat{C}(G) \triangleq \arg \min_{x \in G} \max_{y \in G} \frac{1}{300} d(x, y) P(y)300
$$

by Lemma 1.7 (page 19)

$$
\hat{C}_b(G) \triangleq \arg \min_{x \in G} \max_{y \in G} \sum_{y \in G} d(x, y) P(y)300
$$

by definition of $\hat{C}_b$ (Definition 2.4 page 28)

$$
\hat{C}(G) \triangleq \arg \min_{x \in G} \max_{y \in G} d(x, y) P(y)
$$

by definition of $G$

$$
\hat{C}_b(G) \triangleq \arg \min_{x \in G} \sum_{y \in G} d(x, y) P(y)300
$$

by Lemma 1.2 (page 17)

$$
\hat{C}_b(G) \triangleq \arg \min_{x \in G} \max_{y \in G} \sum_{y \in G} d(x, y) P(y)300
$$

by definition of $\hat{C}_b$ (Definition 2.4 page 28)
\[
\hat{h}_n(G) = \arg \min_{x \in \mathcal{G}} \sum_{y \in \mathcal{Q}(x)} \frac{1}{d(x, y)} P(y) \\
= \arg \max_{x \in \mathcal{G}} \sum_{y \in \mathcal{Q}(x)} \frac{1}{d(x, y)} P(y) \\
= \arg \max_{x \in \mathcal{G}} \frac{1}{300} \sum_{y \in \mathcal{Q}(x)} \frac{1}{d(x, y)} 300P(y) \\
= \arg \max_{x \in \mathcal{G}} \frac{300P(y)}{d(x, y)} \\
= \arg \max_{x \in \mathcal{G}} \left[ \begin{array}{cccccc} 0 & 30 & 15 & 10 & 180 & 6 \\
30 & 0 & 15 & 0 & 180 & 12 & 10 \\
30 & 15 & 0 & 0 & 6 & 10 \\
30 & 30 & 10 & 180 & 0 & 10 \\
60 & 15 & 10 & 180 & 6 & 0 \\
\end{array} \right] \\
= \arg \min_{x \in \mathcal{G}} \left[ \begin{array}{cccccc} 6 \\
10 \\
6 \\
10 \\
6 \\
\end{array} \right] = \{\square, \square\} \\
\]
\[
\hat{M}(G) = \arg \max_{x \in \mathcal{G}} \min_{y \in \mathcal{Q}(x)} \frac{d(x, y)}{P(y)} \\
= \arg \max \{6, 10, 6, 10, 6\} = \{\square, \square\} \\
\]
\[
\hat{V}_\varphi(G) = \sum_{x \in \mathcal{G}} d^2(\hat{h}(G), x) P(x) \\
= \sum_{x \in \mathcal{G}} d^2(\square, x) P(x) \\
= 1^2 \times \frac{1}{10} + 1^2 \times \frac{1}{20} + 2^2 \times \frac{1}{30} + 0^2 \times \frac{3}{5} + 2^2 \times \frac{1}{50} + 1^2 \times \frac{1}{30} \\
= \frac{1}{300} (30 + 15 + 40 + 0 + 6 + 10) = 101 \approx 0.337 \\
\]
\[
\hat{V}_\varphi(G; \hat{c}_k) = \sum_{x \in \mathcal{G}} d^2(\hat{h}_k(G), x) P(x) \\
= \sum_{x \in \mathcal{G}} d^2(\square, x) P(x) \\
= \{\square, \square\} \\
\text{by definition of } \hat{V}_\varphi \\
\]
\[ \begin{align*}
&= 1^2 \times \frac{1}{10} + 0^2 \times \frac{1}{20} + 1^2 \times \frac{1}{30} + 1^2 \times \frac{3}{5} + 2^2 \times \frac{1}{50} + 1^2 \times \frac{1}{30} \\
&= \frac{1}{300} (30 + 0 + 10 + 180 + 24 + 10) = \frac{254}{300} = \frac{127}{150} \approx 0.847 \\
\text{Var}(G; \hat{\xi}_n) &\triangleq \sum_{x \in G} d^2 \left( \hat{\xi}_n(G), x \right) P(x) \\
&= \sum_{x \in G} d^2(\Box, x) P(x) \\
&= 0^2 \times \frac{1}{10} + 1^2 \times \frac{1}{20} + 0^2 \times \frac{1}{30} + 0^2 \times \frac{3}{5} + 1^2 \times \frac{1}{50} + 0^2 \times \frac{1}{30} \\
&= \frac{1}{300} (0 + 60 + 0 + 0 + 6 + 0) = \frac{66}{300} = \frac{11}{50} = 0.22 \\
\text{Var}(G; \hat{\psi}_m) &\triangleq \sum_{x \in G} d^2 \left( \hat{\psi}_m(G), x \right) P(x) \\
&= \sum_{x \in G} d^2(\Box, x) P(x) \\
&= 1^2 \times \frac{1}{10} + 0^2 \times \frac{1}{20} + 1^2 \times \frac{1}{30} + 1^2 \times \frac{3}{5} + 0^2 \times \frac{1}{50} + 1^2 \times \frac{1}{30} \\
&= \frac{1}{300} (30 + 0 + 10 + 180 + 0 + 10) = \frac{230}{300} = \frac{11}{50} \approx 0.767
\end{align*} \]

**Example 2.4** (board game spinner outcome subspace). The six value spinner outcome subspace (Definition 2.10 page 29) has the following geometric values:

\[ \hat{\xi}(G) = \hat{\xi}(G) = \hat{\xi}(G) = \hat{\xi}(G) = \hat{\xi}(G) = \hat{\xi}(G) = \{1, 2, 3, 4, 5, 6\} \]

\[ \text{Var}(G) = \text{Var}(G; \hat{\xi}) = \text{Var}(G; \hat{\psi}) = \text{Var}(G; \hat{\psi}) = \text{Var}(G; \hat{\psi}) = \text{Var}(G; \hat{\psi}) = 0 \]

**Proof:**

\[ \hat{\xi}(G) \triangleq \arg \min_{x \in G} \max_{y \in G} d(x, y) P(y) \quad \text{by definition of } \hat{\xi} \text{ (Definition 2.3 page 28)} \]

\[ = \arg \min_{x \in G} \max_{y \in G} d(x, y) \frac{1}{6} \quad \text{by definition of } G \]

\[ = \arg \min_{x \in G} \max_{y \in G} d(x, y) \quad \text{because } f(x) = \frac{1}{6} \text{ is strictly isotone and by Lemma 1.7 (page 19)} \]

\[ = \\begin{bmatrix}
0 & 1 & 2 & 3 & 2 & 1 \\
1 & 0 & 1 & 2 & 3 & 2 \\
2 & 1 & 0 & 1 & 2 & 3 \\
3 & 2 & 1 & 0 & 1 & 2 \\
2 & 3 & 2 & 1 & 0 & 1 \\
1 & 2 & 3 & 2 & 1 & 0
\end{bmatrix} \quad = \arg \min_{x \in G} \left\{ \begin{bmatrix} 3 \\ 3 \\ 3 \\ 3 \\ 3 \end{bmatrix} \right\} \]

\[ = \{ \Box, \Box, \Box, \Box, \Box, \Box \} \quad \text{by definition of } G \]

\[ \hat{\psi}(G) \triangleq \arg \min_{x \in G} \sum_{y \in G} d(x, y) P(y) \quad \text{by definition of } \hat{\psi} \text{ (Definition 2.4 page 28)} \]

\[ = \arg \min_{x \in G} \sum_{y \in G} d(x, y) \frac{1}{6} \quad \text{by definition of } G \]
\[ = \arg \min_{x \in G} \sum_{y \in G} d(x, y) \quad \text{because } f(x) = \frac{1}{x} \text{ is strictly isotive and by Lemma 1.2 (page 17)} \]

\[ = \arg \min_{x \in G} \begin{cases} 
0 + 1 + 2 + 3 + 2 + 1 \\
1 + 0 + 1 + 2 + 3 + 2 \\
3 + 2 + 1 + 0 + 1 + 2 \\
2 + 2 + 2 + 1 + 0 + 1 \\
1 + 2 + 3 + 2 + 1 + 0
\end{cases} \begin{cases} 
9 \\
9 \\
9 \\
9 \\
9
\end{cases} = \arg \min_{x \in G} \begin{cases} 
9 \\
9 \\
9 \\
9 \\
9
\end{cases} = \begin{cases} 
\text{ } \\
\text{ } \\
\text{ } \\
\text{ } \\
\text{ }
\end{cases} \]

\[ \hat{c}_g(G) \triangleq \arg \min_{x \in G} \prod_{y \in O \backslash \{x\}} [d(x, y)^{p(y)}] \quad \text{by definition of } \hat{c}_g \text{ (Definition 2.4 page 28)} \]

\[ = \arg \min_{x \in G} \prod_{y \in O \backslash \{x\}} d(x, y)^{1} \quad \text{by definition of } G \]

\[ = \arg \min_{x \in G} \left( \prod_{y \in O \backslash \{x\}} d(x, y) \right)^{\frac{1}{g}} \quad \text{by Lemma 1.2 (page 17)} \]

\[ = \arg \min_{x \in G} \begin{cases} 
\times \times 1 \times 2 \times 3 \times 2 \times 1 \\
1 \times 1 \times 1 \times 1 \times 1 \times 1 \\
2 \times 1 \times 1 \times 1 \times 1 \times 1 \\
3 \times 2 \times 1 \times 1 \times 1 \times 1 \\
2 \times 1 \times 1 \times 1 \times 1 \times 1 \\
1 \times 1 \times 1 \times 1 \times 1 \times 1
\end{cases} \begin{cases} 
12 \\
12 \\
12 \\
12 \\
12 \\
12
\end{cases} = \arg \min_{x \in G} \begin{cases} 
12 \\
12 \\
12 \\
12 \\
12 \\
12
\end{cases} = \begin{cases} 
\text{ } \\
\text{ } \\
\text{ } \\
\text{ } \\
\text{ } \\
\text{ }
\end{cases} \]

\[ \hat{c}_n(G) \triangleq \arg \min_{x \in G} \left( \sum_{y \in O \backslash \{x\}} \frac{1}{d(x, y)} P(y) \right)^{-1} \quad \text{by definition of } \hat{c}_n \text{ (Definition 2.4 page 28)} \]

\[ = \arg \max_{x \in G} \sum_{y \in O \backslash \{x\}} \frac{1}{d(x, y)} P(y) \quad \text{because } \phi(x) \triangleq x^{-1} \text{ is strictly antitone and by Lemma 1.5 page 18} \]

\[ = \arg \max_{x \in G} \frac{1}{d(x, y)} \begin{cases} 
1 \\
1 \\
1 \\
1 \\
1
\end{cases} \begin{cases} 
6 \\
6 \\
6 \\
6 \\
6
\end{cases} = \arg \max_{x \in G} \begin{cases} 
6 \\
6 \\
6 \\
6 \\
6
\end{cases} = \begin{cases} 
\text{ } \\
\text{ } \\
\text{ } \\
\text{ } \\
\text{ }
\end{cases} \]

\[ \hat{c}_m(G) \triangleq \arg \min_{x \in G} \min_{y \in O \backslash \{x\}} d(x, y) P(y) \quad \text{by definition of } \hat{c}_m \text{ (Definition 2.4 page 28)} \]

\[ = \arg \min_{x \in G} \min_{y \in O \backslash \{x\}} d(x, y) \frac{1}{6} \]

\[ = \arg \min_{y \in O \backslash \{x\}} d(x, y) \quad \text{by definition of } G \]

\[ = \arg \min_{y \in O \backslash \{x\}} \{1, 1, 1, 1, 1\} \]

\[ = \{\square, \square, \square, \square, \square\} \quad \text{by definition of } G \]

\[ \hat{c}_m(G) \triangleq \arg \max_{x \in G} \min_{y \in O \backslash \{x\}} d(x, y) P(y) \quad \text{by definition of } \hat{c}_m \text{ (Definition 2.4 page 28)} \]

\[ = \arg \max_{x \in G} \{1, 1, 1, 1, 1\} \]

\[ = \{\square, \square, \square, \square, \square\} \quad \text{by } \hat{c}_m(G) \text{ result} \]
= \{ \varnothing, \{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}, \{1, 2, 3\} \} \quad \text{by definition of } G

\textbf{Example 2.5 (weighted spinner outcome subspace).} \\

The six value \textit{weighted spinner outcome subspace} $G$ (Definition 2.1 page 27) illustrated to the right has the following geometric values:

$\hat{\mathcal{C}}(G) = \hat{\mathcal{C}}^1(G) = \hat{\mathcal{C}}^2(G) = \{1, 6\} \quad \text{Var}(G) = \frac{5}{3} \approx 1.667$  

$\hat{\mathcal{C}}^4(G) = \{2, 5\} \quad \text{Var}(G; \hat{\mathcal{C}}^4) = \frac{4}{3} \approx 1.333$  

$\hat{\mathcal{C}}^m(G) = \hat{\mathcal{C}}^m(M) = \{1, 2, 3, 4, 5, 6\} \quad \text{Var}(G; \hat{\mathcal{C}}^m) = 0 = 0$

The \textit{outcome center} result is used later in Example 2.16 (page 57). Note that, unlike the \textit{weighted real die outcome subspace} (Example 2.3 page 33), of the center measures of cardinality 2 or less, the \textit{harmonic center} $\hat{\mathcal{C}}^h(G)$ yields the lowest \textit{outcome variance} (Definition 2.5 page 28). This is surprising since it suggests that $\hat{\mathcal{C}}^h(G)$ is superior to all the other \textit{center measures} (Definition 2.3 page 28, Definition 2.4 page 28), but yet unlike the other center measures, it yields center values that are not maximally likely.

\hspace{1cm} \textbf{Proof:}

$\hat{\mathcal{C}}(G) \triangleq \arg \min_{x \in G} \max_{y \in G} \frac{1}{10} \sum_{x \in G} d(x, y) P(y)$ \quad \text{by definition of } \hat{\mathcal{C}} (Definition 2.3 page 28)

$= \arg \min_{x \in G} \max_{y \in G} \frac{1}{10} \{ 0 \times 3 \times 1 \times 1 \times 2 \times 1 \times 3 \times 1 \times 2 \times 1 \times 1 \times 3 \}$

$= \arg \min_{x \in G} \max_{y \in G} \frac{1}{10} \{ 1 \times 3 \times 0 \times 1 \times 1 \times 2 \times 1 \times 3 \times 1 \times 2 \times 3 \}$

$= \arg \min_{x \in G} \max_{y \in G} \frac{1}{10} \{ 2 \times 3 \times 3 \times 1 \times 1 \times 1 \times 1 \times 2 \times 3 \}$

$= \arg \min_{x \in G} \max_{y \in G} \frac{1}{10} \{ 3 \times 3 \times 3 \times 1 \times 1 \times 1 \times 1 \times 2 \times 3 \}$

$= 1 \quad \{ 3 \quad 6 \quad 9 \}$

$= \{ 6 \}$

$\hat{\mathcal{C}}^h(G) \triangleq \arg \min_{x \in G} \prod_{y \in G(x)} d(x, y)^{p(y)} \quad \text{by definition of } \hat{\mathcal{C}}^h (Definition 2.4 page 28)$

$= \arg \min_{x \in G} \prod_{y \in G(x)} \left[ d(x, y)^{p(y)} \right]^{\frac{1}{2}}$

$= \arg \min_{x \in G} \left( \prod_{y \in G(x)} \left[ d(x, y)^{p(y)} \right] \right)^{\frac{1}{2}}$

$= \arg \min_{x \in G} \prod_{y \in G(x)} \left[ d(x, y)^{p(y)} \right] \quad \text{by Lemma 1.2 (page 17)}$
\[
\hat{L}_{\infty}(G) \triangleq \arg \min_{x \in G} \left\{ \sum_{y \in \Omega \mid y(x)} \frac{1}{d(x, y)} P(y) \right\}^{-1}
\]
by definition of \( \hat{L}_{\infty} \) (Definition 2.4 page 28)

\[
= \arg \max_{x \in G} \sum_{y \in \Omega \mid y(x)} \frac{1}{d(x, y)} P(y)
\]
because \( \phi(x) \triangleq x^{-1} \) is strictly antitone and by Lemma 1.5 page 18

\[
= \arg \max_{x \in G} \sum_{y \in \Omega \mid y(x)} \frac{1}{d(x, y)} P(y) \frac{1}{6}
\]

\[
= \arg \max_{x \in G} \sum_{y \in \Omega \mid y(x)} 6P(y) \frac{1}{d(x, y)}
\]
by Lemma 1.2 (page 17)

\[
= \arg \max_{x \in G} \left\{ \frac{\frac{1}{d} + \frac{1}{x} + \frac{1}{y} + \frac{1}{z} + \frac{1}{w} + \frac{1}{v} + \frac{1}{u} + \frac{1}{t} + \frac{1}{s} + \frac{1}{r} + \frac{1}{q} + \frac{1}{p}}{3} \right\}
\]

\[
= \arg \max_{x \in G} \left\{ \frac{1}{6} \right\} \left\{ \begin{array}{c} 32 \\ 38 \\ 30 \\ 30 \\ 38 \\ 32 \end{array} \right\} = \left\{ \begin{array}{c} 2 \\ 5 \end{array} \right\}
\]

\[
\hat{L}_{m}(G) \triangleq \arg \min_{x \in G} \min_{y \in \Omega \mid y(x)} \left\{ \begin{array}{c} 3 \\ 6 \\ 9 \\ 6 \\ 3 \\ 2 \end{array} \right\}
\]
by definition of \( \hat{L}_{m} \) (Definition 2.4 page 28)

\[
= \arg \max_{x \in G} \{1, 1, 1, 1, 1, 1\}
\]
by \( \hat{L}_{m}(G) \) result

\[
= \{1, 2, 3, 4, 5, 6\}
\]

\[
\hat{\text{Var}}(G; \hat{L}_{m}) = \text{Var}(G; \hat{L}_{m}) = \text{Var}(G)
\]
by definition of \( \hat{\text{Var}} \) (Definition 2.5 page 28)

\[
\triangleq \sum_{x \in G} d^2\left( \hat{L}(G), x \right) P(x)
\]
by \( \hat{L}(G) \) result

\[
= \sum_{x \in G} d^2\left( (1, 6), x \right) P(x)
\]

\[
= (0)^2 \frac{3}{6} + (1)^2 \frac{1}{6} + (2)^2 \frac{1}{6} + (2)^2 \frac{1}{6} + (1)^2 \frac{1}{6} + (0)^2 \frac{3}{6}
\]

\[
= \frac{10}{3} = \frac{5}{3} = \frac{5}{3} \approx 1.667
\]

\[
\text{Var}(G; \hat{L}_{m}) \triangleq \sum_{x \in G} d^2\left( \hat{L}_{m}(G), x \right) P(x)
\]
by \( \hat{L}_{m}(G) \) result

\[
= \sum_{x \in G} d^2\left( (2, 5), x \right) P(x)
\]
by \( \hat{L}_{m}(G) \) result

\[
= (1)^2 \frac{3}{6} + (0)^2 \frac{1}{6} + (1)^2 \frac{1}{6} + (1)^2 \frac{1}{6} + (0)^2 \frac{1}{6} + (1)^2 \frac{3}{6}
\]

\[
= \frac{8}{6} = \frac{4}{3} \approx 1.333
\]

\[
\text{Var}(G; \hat{L}_{m}) = \text{Var}(G; \hat{L}_{m})
\]
\[ \Delta \sum_{x \in G} d^2 \left( \hat{\zeta}_n(G), x \right) P(x) \quad \text{by definition of } \var{ \text{Var} } \text{ (Definition 2.5 page 28)} \]
\[ = \sum_{x \in G} d^2((1,2,3,4,5,6), x) P(x) \quad \text{by } \hat{\zeta}_n(G) \text{ result} \]
\[ = \sum_{x \in G} 0^2 P(x) = 0 \]

\[ \text{Example 2.6 (weighted ring). The weighted five element ring illustrated to the right has the geometric values below. The outcome center result is used later in Example 2.17 (page 59).} \]
\[ \hat{\zeta}(G) = \{4\} \quad \var{ \text{Var} } (G) = \frac{11}{9} \approx 1.222 \]
\[ \hat{\zeta}_2(G) = \{3,4\} \quad \var{ \text{Var} } (G; \hat{\zeta}_2) = \frac{7}{9} \approx 0.778 \]
\[ \hat{\zeta}_3(G) = \{3\} \quad \var{ \text{Var} } (G; \hat{\zeta}_3) = \frac{16}{9} \approx 1.778 \]
\[ \hat{\zeta}_4(G) = \{1,2,3\} \quad \var{ \text{Var} } (G; \hat{\zeta}_4) = \frac{5}{9} \approx 0.556 \]
\[ \hat{\zeta}_5(G) = \{0,3,4\} \quad \var{ \text{Var} } (G; \hat{\zeta}_5) = \frac{2}{9} \approx 0.222 \]
\[ \hat{\zeta}_M(G) = \{1,2\} \quad \var{ \text{Var} } (G; \hat{\zeta}_M) = \frac{16}{9} \approx 1.778 \]

\[ \text{Proof:} \]
\[ \hat{\zeta}(G) \triangleq \arg \min_{x \in G} \max_{y \in G} d(x, y) P(y) \quad \text{by definition of } \hat{\zeta} \text{ (Definition 2.3 page 28)} \]
\[ = \arg \min_{x \in G} \max_{y \in G} \frac{1}{9} d(x, y) P(y) \]
\[ = \arg \min_{x \in G} \max_{y \in G} d(x, y) P(y) \quad \text{because } f(x) = \frac{1}{9} x \text{ is strictly isotone and by Lemma 1.7 (page 19)} \]
\[ = \arg \min_{x \in G} \max_{y \in G} \left\{ \begin{array}{c}
0 \times 2 + 1 \times 1 + 2 \times 2 + 1 \times 3 \\
1 \times 2 + 0 \times 1 + 1 \times 1 + 2 \times 2 + 2 \times 3 \\
2 \times 2 + 1 \times 1 + 0 \times 1 + 1 \times 2 + 2 \times 3 \\
1 \times 2 + 2 \times 1 + 2 \times 1 + 1 \times 2 + 0 \times 3 \\
\end{array} \right\} = \arg \min_{x \in G} \left\{ \begin{array}{c}
4 \\
6 \\
6 \\
4 \\
\end{array} \right\} \]
\[ \hat{\zeta}_y(G) \triangleq \arg \min_{x \in G} \sum_{y \in G} d(x, y) P(y) \quad \text{by definition of } \hat{\zeta}_y \text{ (Definition 2.4 page 28)} \]
\[ = \arg \min_{x \in G} \sum_{y \in G} \frac{1}{9} d(x, y) P(y) \]
\[ = \arg \min_{x \in G} \sum_{y \in G} d(x, y) P(y) \quad \text{because } f(x) = \frac{1}{9} x \text{ is strictly isotone and by Lemma 1.2 (page 17)} \]
\[ = \arg \min_{x \in G} \left\{ \begin{array}{c}
0 \times 2 + 2 \times 1 + 1 \times 1 + 1 \times 2 + 2 \times 3 \\
2 \times 2 + 0 \times 1 + 2 \times 1 + 1 \times 2 + 1 \times 3 \\
1 \times 2 + 2 \times 1 + 0 \times 1 + 2 \times 2 + 1 \times 3 \\
2 \times 2 + 1 \times 1 + 2 \times 1 + 1 \times 2 + 0 \times 3 \\
\end{array} \right\} = \arg \min_{x \in G} \left\{ \begin{array}{c}
11 \\
11 \\
8 \\
8 \\
\end{array} \right\} = \arg \min_{x \in G} \left\{ \begin{array}{c}
3 \\
4 \\
\end{array} \right\} \]
\[ \hat{\zeta}_k(G) \triangleq \arg \min_{x \in G} \prod_{y \in G \setminus \{x\}} d(x, y)^{P(y)} \quad \text{by definition of } \hat{\zeta}_k \text{ (Definition 2.4 page 28)} \]
\[ = \arg \min_{x \in G} \prod_{y \in G \setminus \{x\}} d(x, y)^{P(y)^{1/2}} \]
\[ = \arg \min_{x \in G} \left[ \prod_{y \in G \setminus \{x\}} d(x, y)^{P(y)} \right]^{1/2} \quad \text{because } f(x) \triangleq x^{1/2} \text{ is strictly isotone and by Lemma 1.2 (page 17)} \]
\[
\hat{\ell}_n(G) = \arg \min_{x \in G} \left\{ \sum_{y \in \Omega} \frac{1}{d(x, y)} P(y) \right\}^{-1}
\]
by definition of \(\hat{\ell}_n\) (Definition 2.4 page 28)

\[
\hat{\ell}_m(G) = \arg \max_{x \in G} \min_{y \in \Omega_{x \mid x}} d(x, y) P(y)
\]
by definition of \(\hat{\ell}_m\) (Definition 2.4 page 28)

\[
\text{Var}(G) = \sum_{x \in G} d^2(\hat{\ell}(G), x) P(x)
\]
by definition of \(\text{Var}\) (Definition 2.5 page 28)

\[
\text{Var}(G; \hat{\ell}_n) = \sum_{x \in G} d^2(\hat{\ell}_n(G), x) P(x)
\]
by definition of \(\text{Var}\) (Definition 2.5 page 28)

\[
\text{Var}(G; \hat{\ell}_m) = \sum_{x \in G} d^2(\hat{\ell}_m(G), x) P(x)
\]
by definition of \(\text{Var}\) (Definition 2.5 page 28)
\( = (2)^2 \frac{2}{9} + (2)^2 \frac{1}{9} + (1)^2 \frac{1}{9} + (0)^2 \frac{2}{9} + (1)^2 \frac{3}{9} = \frac{16}{9} \approx 1.778 \)

\[ \text{Var}(\mathcal{G}; \hat{\mathcal{C}}_k) \triangleq \sum_{x \in \mathcal{G}} d^2(\hat{C}_k(\mathcal{G}), x) P(x) \quad \text{by definition of \text{Var} (Definition 2.5 page 28)} \]

\[ = \sum_{x \in \mathcal{G}} d^2((1, 2, 3), x) P(x) \quad \text{by \hat{C}_k(\mathcal{G}) result} \]

\[ = (1)^2 \frac{2}{9} + (0)^2 \frac{1}{9} + (0)^2 \frac{1}{9} + (0)^2 \frac{2}{9} + (1)^2 \frac{3}{9} = \frac{5}{9} \approx 0.556 \]

\[ \text{Var}(\mathcal{G}; \hat{\mathcal{C}}_m) \triangleq \sum_{x \in \mathcal{G}} d^2(\hat{C}_m(\mathcal{G}), x) P(x) \quad \text{by definition of \text{Var} (Definition 2.5 page 28)} \]

\[ = \sum_{x \in \mathcal{G}} d^2((0, 3, 4), x) P(x) \quad \text{by \hat{C}_m(\mathcal{G}) result} \]

\[ = (0)^2 \frac{2}{9} + (1)^2 \frac{1}{9} + (1)^2 \frac{1}{9} + (0)^2 \frac{2}{9} + (0)^2 \frac{3}{9} = \frac{2}{9} \approx 0.222 \]

\[ \text{Var}(\mathcal{G}; \hat{\mathcal{C}}_m) \triangleq \sum_{x \in \mathcal{G}} d^2(\hat{C}_m(\mathcal{G}), x) P(x) \quad \text{by definition of \text{Var} (Definition 2.5 page 28)} \]

\[ = \sum_{x \in \mathcal{G}} d^2((1, 2), x) P(x) \quad \text{by \hat{C}_m(\mathcal{G}) result} \]

\[ = (1)^2 \frac{2}{9} + (0)^2 \frac{1}{9} + (0)^2 \frac{1}{9} + (1)^2 \frac{2}{9} + (2)^2 \frac{3}{9} = \frac{16}{9} \approx 1.778 \]

\hspace{1cm} \text{Example 2.7.} \text{ The weighted five element structure illustrated to the right has the following geometric values:} 
\[ \hat{C}(\mathcal{G}) = \hat{C}_k(\mathcal{G}) = \{3\} \]
\[ \hat{C}_k(\mathcal{G}) = \{3, 4\} \quad \text{Var}(\mathcal{G}) = \frac{10}{9} \approx 1.111 \]
\[ \hat{C}_m(\mathcal{G}) = \{1, 2, 3\} \quad \text{Var}(\mathcal{G}; \hat{C}_k) = \frac{5}{9} \approx 0.555 \]
\[ \hat{C}_m(\mathcal{G}) = \{0, 3, 4\} \quad \text{Var}(\mathcal{G}; \hat{C}_m) = \frac{2}{9} \approx 0.222 \]
\[ \hat{C}_m(\mathcal{G}) = \{1, 2\} \quad \text{Var}(\mathcal{G}; \hat{C}_m) = \frac{7}{9} \approx 0.778 \]
\hspace{1cm} \text{The outcome center result is used later in Example 2.18 (page 61). Note that only the operators \hat{C} and \hat{C}_k were able to successfully isolate a single center point (|\hat{C}(\mathcal{G})| = |\hat{C}_k(\mathcal{G})| = |\{3\}| = 1).} 
\hspace{1cm} \text{\textcopyright\textsc{proof:}} 
\[ \hat{C}(\mathcal{G}) \triangleq \arg\min_{x \in \mathcal{G}} \arg\min_{y \in \mathcal{G}} d(x, y) P(y) \quad \text{by definition of \hat{C} (Definition 2.3 page 28)} \]
\[ = \arg\min_{x \in \mathcal{G}} \arg\min_{y \in \mathcal{G}} \frac{1}{2} d(x, y) P(y) \]
\[ = \arg\min_{x \in \mathcal{G}} \arg\min_{y \in \mathcal{G}} d(x, y) P(y) \quad \text{because } f(x) = \frac{1}{9} x \text{ is strictly isotone and by Lemma 1.7 (page 19)} \]
\[ = \arg\min_{x \in \mathcal{G}} \arg\min_{y \in \mathcal{G}} \begin{cases} d(0, 0) P(0) & d(0, 1) P(1) & d(0, 2) P(2) & d(0, 3) P(3) & d(0, 4) P(4) \\ d(1, 0) P(0) & d(1, 1) P(1) & d(1, 2) P(2) & d(1, 3) P(3) & d(1, 4) P(4) \\ d(2, 0) P(0) & d(2, 1) P(1) & d(2, 2) P(2) & d(2, 3) P(3) & d(2, 4) P(4) \\ d(3, 0) P(0) & d(3, 1) P(1) & d(3, 2) P(2) & d(3, 3) P(3) & d(3, 4) P(4) \\ d(4, 0) P(0) & d(4, 1) P(1) & d(4, 2) P(2) & d(4, 3) P(3) & d(4, 4) P(4) \\ d(4, 3) P(3) & d(4, 4) P(4) & d(3, 4) P(4) & d(2, 4) P(4) & d(1, 4) P(4) \\ d(2, 3) P(3) & d(1, 3) P(3) & d(0, 3) P(3) & d(0, 4) P(4) & d(1, 4) P(4) \\ d(1, 2) P(2) & d(0, 2) P(2) & d(0, 1) P(1) & d(0, 0) P(0) & d(1, 0) P(0) \\ d(2, 1) P(1) & d(1, 1) P(1) & d(1, 0) P(0) & d(1, 0) P(0) & d(0, 0) P(0) \\ d(3, 1) P(1) & d(2, 1) P(1) & d(1, 1) P(1) & d(0, 1) P(1) & d(0, 0) P(0) \end{cases} \]
\[ = \arg\min_{x \in \mathcal{G}} \begin{cases} 6 \quad 4 \\ 4 \quad 3 \\ 3 \quad 4 \end{cases} \]
\[ \hat{C}_k(\mathcal{G}) \triangleq \arg\min_{x \in \mathcal{G}} \sum_{y \in \mathcal{G}} d(x, y) P(y) \quad \text{by definition of \hat{C}_k (Definition 2.4 page 28)} \]
\[ = \arg\min_{x \in \mathcal{G}} \begin{cases} 0 \times 2 + 2 \times 1 + 1 \times 1 + 1 \times 2 + 2 \times 3 \\ 2 \times 2 + 0 \times 1 + 2 \times 1 + 1 \times 2 + 1 \times 3 \\ 1 \times 2 + 2 \times 1 + 0 \times 1 + 2 \times 2 + 1 \times 3 \\ 1 \times 2 + 1 \times 1 + 2 \times 1 + 0 \times 2 + 1 \times 3 \\ 2 \times 2 + 1 \times 1 + 1 \times 1 + 1 \times 2 + 0 \times 3 \end{cases} \]
\[ = \arg\min_{x \in \mathcal{G}} \begin{cases} 11 \\ 11 \\ 8 \\ 8 \\ 3 \end{cases} \]
\[ \hat{\ell}_G(G) \triangleq \arg \min_{x \in G} \prod_{y \in G(x)} d(x, y)^{P(y)} \] 
by definition of \( \hat{\ell}_G \) (Definition 2.4 page 28)

\[ \begin{align*}
= \arg \min_{x \in G} \prod_{y \in G(x)} d(x, y)^{q_P(y)}
= \arg \min_{x \in G} \left[ \prod_{y \in G(x)} d(x, y)^{q_P(y)} \right]^{1/2}
= \arg \min_{x \in G} \prod_{y \in G(x)} d(x, y)^{q_P(y)}
\end{align*} \]
because \( f(x) \triangleq x^{1/2} \) is strictly monotone and by Lemma 1.2 (page 17)

\[ \hat{\ell}_\Omega(G) \triangleq \arg \min_{x \in \Omega} \left( \sum_{y \in \Omega} \frac{1}{d(x, y)} P(y) \right)^{-1} \] 
by definition of \( \hat{\ell}_\Omega \) (Definition 2.4 page 28)

\[ \begin{align*}
= \arg \max_{x \in \Omega} \left( \sum_{y \in \Omega} \frac{1}{d(x, y)} P(y) \right)
= \arg \max_{x \in \Omega} \left( \frac{1}{9} \sum_{y \in \Omega} \frac{1}{d(x, y)} P(y) \right)^9
= \arg \max_{x \in \Omega} \frac{9P(y)}{d(x, y)}
\end{align*} \]
because \( f(x) = \frac{1}{9} x \) is strictly monotone and by Lemma 1.2 (page 17)

\[ \hat{\ell}_m(G) \triangleq \arg \min_{x \in G} \prod_{y \in G(x)} d(x, y)^{P(y)} \] 
by definition of \( \hat{\ell}_m \) (Definition 2.4 page 28)

\[ \begin{align*}
= \arg \min_{x \in G} \prod_{y \in G(x)} d(x, y)^{q_P(y)}
= \arg \min_{x \in G} \left[ \prod_{y \in G(x)} d(x, y)^{q_P(y)} \right]^{1/9}
= \arg \min_{x \in G} \prod_{y \in G(x)} d(x, y)^{q_P(y)}
\end{align*} \]
because \( f(x) = \frac{1}{9} x \) is strictly monotone and by Lemma 1.7 (page 19)

\[ \hat{\ell}_M(G) \triangleq \arg \max_{x \in G} \min_{y \in G(x)} d(x, y)^{P(y)} \] 
by definition of \( \hat{\ell}_M \) (Definition 2.4 page 28)

\[ \begin{align*}
= \arg \max_{x \in G} \{ 1, 2, 2, 1, 1 \}
= \{ 1, 2 \}
\end{align*} \]
\( \hat{\ell}_M(G) \) result

\[ \mathcal{V}\mathcal{A}(G) = \mathcal{V}\mathcal{A}(G; \hat{\ell}_G) \] 
by \( \hat{\ell}(G) \) and \( \hat{\ell}_G(G) \) results

\[ \begin{align*}
= \sum_{x \in G} d^2(\hat{\ell}_G(G), x) P(x)
= \sum_{x \in G} d^2([3], x) P(x)
= (1)^2 \frac{2}{9} + (1)^2 \frac{2}{9} + (2)^2 \frac{2}{9} + (0)^2 \frac{2}{9} + (1)^2 \frac{3}{9} = \frac{10}{9} \approx 1.111
\]
\[ \text{Var}(G; \hat{\mu}) \triangleq \sum_{x \in G} d^2(\hat{\mu}(G), x) P(x) \quad \text{by definition of } \text{Var} \text{ (Definition 2.5 page 28)} \]
\[ = \sum_{x \in G} d^2([3, 4], x) P(x) \quad \text{by } \hat{\mu}(G) \text{ result} \]
\[ = (1)^2 \frac{2}{9} + (1)^2 \frac{1}{9} + (1)^2 \frac{1}{9} + (0)^2 \frac{2}{9} + (0)^2 \frac{3}{9} = \frac{16}{9} \approx 0.444 \]
\[ \text{Var}(G; \hat{\mu}_n) \triangleq \sum_{x \in G} d^2(\hat{\mu}_n(G), x) P(x) \quad \text{by definition of } \text{Var} \text{ (Definition 2.5 page 28)} \]
\[ = \sum_{x \in G} d^2([1, 2, 3], x) P(x) \quad \text{by } \hat{\mu}_n(G) \text{ result} \]
\[ = (1)^2 \frac{2}{9} + (0)^2 \frac{1}{9} + (0)^2 \frac{1}{9} + (0)^2 \frac{2}{9} + (1)^2 \frac{3}{9} = \frac{7}{9} \approx 0.778 \]

Example 2.8. The outcome subspace (Definition 2.1 page 27) illustrated to the right, with a quasi-metric (Definition D.6 page 153) has the following geometric values:
\[ \hat{\mu}(G) = \hat{\mu}_k(G) = \hat{\mu}_k(G) = \{3\} \quad \text{Var}(G) = \frac{12}{9} \approx 1.333 \]
\[ \hat{\mu}_n(G) = \{0, 4\} \quad \text{Var}(G; \hat{\mu}_n) = \frac{10}{9} \approx 1.111 \]
\[ \hat{\mu}_m(G) = \{1, 2, 3\} \quad \text{Var}(G; \hat{\mu}_m) = \frac{5}{9} \approx 0.555 \]

This is the first example in this section to use a directed graph (rather than an undirected graph Definition 1.17 page 7) and to require the use of a quasi-metric (Definition D.6 page 153) that is not a metric. Unlike Example 2.7 (page 41), which had neither of these restrictions, twice as many center operators (4 rather than 2) were able to successfully isolate a single center point. The outcome center result is used later in Example 2.19 (page 62).

\[ \triangleq \text{Proof:} \]
\[ \hat{\mu}(G) \triangleq \arg \min_{x \in G} \max_{y \in G} d(x, y) P(y) \quad \text{by definition of } \hat{\mu} \text{ (Definition 2.3 page 28)} \]
\[ \triangleq \arg \min_{x \in G} \max_{y \in G} \frac{1}{9} d(x, y) P(y) 9 \]
\[ \triangleq \arg \min_{x \in G} \max_{y \in G} d(x, y) P(y) 9 \quad \text{because } f(x) = \frac{1}{9} x \text{ is strictly isotone and by Lemma 1.7 (page 19)} \]
\[ = \arg \min_{x \in G} \max_{y \in G} \left\{ \begin{array}{c} d(0, 0) P(0) 9 \quad d(0, 1) P(1) 9 \quad d(0, 2) P(2) 9 \quad d(0, 3) P(3) 9 \quad d(0, 4) P(4) 9 \\ d(1, 0) P(0) 9 \quad d(1, 1) P(1) 9 \quad d(1, 2) P(2) 9 \quad d(1, 3) P(3) 9 \quad d(1, 4) P(4) 9 \end{array} \right\} \]
\[ = \arg \min_{x \in G} \max_{y \in G} \left\{ \begin{array}{c} 0 \times 2 \quad 3 \times 1 \quad 1 \times 1 \quad 4 \times 2 \quad 2 \times 3 \\ 2 \times 0 \quad 0 \times 1 \quad 2 \times 1 \quad 2 \times 2 \quad 2 \times 3 \\ 4 \times 2 \quad 2 \times 1 \quad 0 \times 1 \quad 3 \times 2 \quad 1 \times 2 \quad 1 \times 3 \\ 1 \times 2 \quad 2 \times 1 \quad 2 \times 1 \quad 0 \times 2 \quad 1 \times 2 \quad 1 \times 3 \\ 3 \times 2 \quad 1 \times 1 \quad 4 \times 1 \quad 2 \times 2 \quad 0 \times 3 \end{array} \right\} = \arg \min_{x \in G} \left\{ \begin{array}{c} 8 \quad 6 \\ 8 \quad 3 \\ 6 \quad 3 \end{array} \right\} \]
\[ \hat{C}_n(G) \triangleq \arg \min_{x \in G} \sum_{y \in G} d(x, y) P(y) \quad \text{by definition of } \hat{C}_n \text{ (Definition 2.4 page 28)} \]

\[ = \arg \min_{x \in G} \sum_{y \in G} \frac{1}{9} d(x, y) P(y) \]

\[ = \arg \min_{x \in G} \sum_{y \in G} d(x, y) P(y) \quad \text{because } f(x) = \frac{1}{9} x \text{ is strictly isotone and by Lemma 1.2 (page 17)} \]

\[ = \arg \min_{x \in G} \max_{y \in G} \begin{bmatrix} 0 \times 2 + 3 \times 1 & + 1 \times 1 & + 4 \times 2 & + 2 \times 3 \\ 2 \times 2 & + 0 \times 1 & + 3 \times 1 & + 1 \times 2 & + 2 \times 3 \\ 2 \times 2 & + 2 \times 1 & + 0 \times 1 & + 3 \times 2 & + 1 \times 3 \\ 1 \times 2 & + 2 \times 1 & + 2 \times 1 & + 0 \times 2 & + 1 \times 3 \\ 3 \times 2 & + 1 \times 1 & + 4 \times 1 & + 2 \times 2 & + 0 \times 3 \end{bmatrix} = \arg \min_{x \in G} \begin{bmatrix} 18 \\ 15 \\ 19 \\ 9 \\ 15 \end{bmatrix} = \begin{bmatrix} 3 \end{bmatrix} \]

\[ \hat{C}_k(G) \triangleq \arg \min_{x \in G} \prod_{y \in G_{[x]}} d(x, y) P(y) \quad \text{by definition of } \hat{C}_k \text{ (Definition 2.4 page 28)} \]

\[ = \arg \min_{x \in G} \prod_{y \in G_{[x]}} d(x, y) P(y)^{\frac{1}{2}} \]

\[ = \arg \min_{x \in G} \left( \prod_{y \in G_{[x]}} d(x, y) P(y) \right)^{\frac{1}{2}} \]

\[ = \arg \min_{x \in G} \prod_{y \in G_{[x]}} d(x, y) P(y) \quad \text{because } f(x) \triangleq x^{\frac{1}{2}} \text{ is strictly isotone and by Lemma 1.2 (page 17)} \]

\[ = \arg \min_{x \in G} \begin{bmatrix} 2^2 & \times & 3^1 & \times & 1^1 & \times & 4^2 & \times & 2^3 \\ 4^2 & \times & 2^1 & \times & 1^2 & \times & 3^2 & \times & 1^3 \\ 1^2 & \times & 2^1 & \times & 2^1 & \times & 1^3 & \times & 1^3 \\ 3^2 & \times & 1^1 & \times & 4^1 & \times & 2^2 \end{bmatrix} = \arg \min_{x \in G} \begin{bmatrix} 384 \\ 192 \\ 243 \\ 144 \end{bmatrix} = \begin{bmatrix} 3 \end{bmatrix} \]

\[ \hat{C}_n(G) \triangleq \arg \min_{x \in \Omega} \left( \sum_{y \in \Omega} \frac{1}{d(x, y)} P(y) \right)^{-1} \quad \text{by definition of } \hat{C}_n \text{ (Definition 2.4 page 28)} \]

\[ = \arg \max_{x \in \Omega} \left( \sum_{y \in \Omega} \frac{1}{d(x, y)} P(y) \right) \quad \text{because } \phi(x) \triangleq x^{-1} \text{ is strictly antitone and by Lemma 1.5 page 18} \]

\[ = \arg \max_{x \in \Omega} \left( \sum_{y \in \Omega} \frac{1}{d(x, y)} P(y)^{\frac{1}{9}} \right) \]

\[ = \arg \max_{x \in \Omega} \sum_{y \in \Omega} \frac{P(y)}{d(x, y)} \quad \text{because } f(x) = \frac{1}{9} x \text{ is strictly isotone and by Lemma 1.2 (page 17)} \]

\[ = \arg \max_{x \in G} \begin{bmatrix} 0 & + & \frac{1}{3} & + & \frac{1}{3} & + & \frac{2}{3} & + & \frac{3}{3} \\ 2 & + & 0 & + & \frac{1}{3} & + & \frac{1}{3} & + & \frac{2}{3} \\ 3 & + & 0 & + & \frac{1}{3} & + & \frac{1}{3} & + & \frac{2}{3} \\ 0 & + & \frac{1}{3} & + & \frac{1}{3} & + & 0 & + & \frac{2}{3} \end{bmatrix} = \arg \max_{x \in G} \begin{bmatrix} 20 \\ 27 \\ 28 \\ 36 \end{bmatrix} = \begin{bmatrix} 3 \end{bmatrix} \]

\[ \hat{C}_n(G) \triangleq \arg \min_{x \in G} \min_{y \in G_{[x]}} d(x, y) P(y) \quad \text{by definition of } \hat{C}_n \text{ (Definition 2.4 page 28)} \]

\[ = \arg \min_{x \in G} \min_{y \in G_{[x]}} \frac{1}{9} d(x, y) P(y)^{\frac{1}{9}} \]

\[ = \arg \min_{x \in G} \min_{y \in G_{[x]}} d(x, y) P(y)^{\frac{1}{9}} \quad \text{because } f(x) = \frac{1}{9} x \text{ is strictly isotone and by Lemma 1.7 (page 19)} \]
\[
\hat{\epsilon}_m(G) = \arg\min_{x \in G} \min_{y \in \partial x} \left\{ \begin{array}{c} 2 \times 2 \\ 4 \times 2 \\ 1 \times 1 \\ 3 \times 2 \\ 2 \times 1 \\ 3 \times 2 \\ 1 \times 1 \\ 4 \times 1 \\ 2 \times 2 \\ 1 \times 3 \\ 2 \times 3 \\ 1 \times 3 \\ 2 \times 3 \\ 1 \times 3 \\ 2 \times 3 \end{array} \right\} = \arg\min_{x \in G} \begin{bmatrix} 1 \\ 2 \\ 2 \\ 1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 0 \\ 4 \end{bmatrix} 
\]

\[
\hat{\epsilon}_m(G) = \arg\max_{x \in G} \\{1, 2, 2, 2, 1\} \text{ by definition of } \hat{\epsilon}_m \text{ (Definition 2.4 page 28)}
\]

\[
\Delta = \arg\max_{x \in G} \\{1, 2, 2, 2, 1\} \text{ by } \hat{\epsilon}_m(G) \text{ result}
\]

\[
\mathbb{V}(G) = \mathbb{V}(\hat{\epsilon}_m(G), \hat{\epsilon}_m(G)) = \mathbb{V}(\hat{\epsilon}_m(G), \hat{\epsilon}_m(G), \hat{\epsilon}_m(G), \hat{\epsilon}_m(G), \hat{\epsilon}_m(G)) \text{ results}
\]

\[
\Delta = \sum_{x \in G} d^2(\hat{\epsilon}_m(G), x) P(x) \text{ by definition of } \mathbb{V} \text{ (Definition 2.5 page 28)}
\]

\[
= \sum_{x \in G} d^2(\{3\}, x) P(x) \text{ by } \hat{\epsilon}_m(G) \text{ result}
\]

\[
= (1)^2 \frac{2}{9} + (2)^2 \frac{1}{9} + (2)^2 \frac{1}{9} + (0)^2 \frac{2}{9} + (1)^2 \frac{3}{9} = \frac{12}{9} = \frac{4}{3} \approx 1.333
\]

\[
\mathbb{V}(G; \hat{\epsilon}_m) = \sum_{x \in G} d^2(\hat{\epsilon}_m(G), x) P(x) \text{ by definition of } \mathbb{V} \text{ (Definition 2.5 page 28)}
\]

\[
= \sum_{x \in G} d^2(\{0, 4\}, x) P(x) \text{ by } \hat{\epsilon}_m(G) \text{ result}
\]

\[
= (0)^2 \frac{2}{9} + (1)^2 \frac{1}{9} + (1)^2 \frac{1}{9} + (2)^2 \frac{2}{9} + (0)^2 \frac{3}{9} = \frac{10}{9} \approx 1.111
\]

\[
\Delta = \sum_{x \in G} d^2(\hat{\epsilon}_m(G), x) P(x) \text{ by definition of } \mathbb{V} \text{ (Definition 2.5 page 28)}
\]

\[
= \sum_{x \in G} d^2(\{1, 2, 3\}, x) P(x) \text{ by } \hat{\epsilon}_m(G) \text{ result}
\]

\[
= (1)^2 \frac{2}{9} + (0)^2 \frac{1}{9} + (0)^2 \frac{1}{9} + (0)^2 \frac{2}{9} + (1)^2 \frac{3}{9} = \frac{5}{9} \approx 0.555
\]

\[
\text{Example 2.9 (DNA). Genomic Signal Processing (GSP) analyzes biological sequences called genomes. These sequences are constructed over a set of 4 symbols that are commonly referred to as } \{\text{\texttt{A}}, \text{\texttt{T}}, \text{\texttt{C}}, \text{\texttt{G}}\}, \text{each of which corresponds to a nucleobase (adenine, thymine, cytosine, and guanine, respectively). A typical genome sequence contains a large number of symbols (about 3 billion for humans, 29751 for the SARS virus). Let } G = \{\{\text{\texttt{A}}, \text{\texttt{T}}, \text{\texttt{C}}, \text{\texttt{G}}\}, \text{\texttt{d}}, \text{\texttt{L}}\text{, } \text{\texttt{P}}\} \text{ be the outcome subspace (Definition 2.1 page 27) generated by a genome where } d \text{ is the discrete metric (Definition D.8 page 155), } \text{\\texttt{d}} \text{ (completely unordered set), and } P(\text{\texttt{A}}) = P(\text{\texttt{T}}) = P(\text{\texttt{C}}) = P(\text{\texttt{G}}) = \frac{1}{4}. \text{ This space is illustrated by the graph (Definition 1.17 page 7) to the right with shaded center (Definition 2.3 page 28). The graph has the following geometric values:}
\]

\[
\hat{\epsilon}_m(G) = \{\{\text{\texttt{A}}, \text{\texttt{T}}, \text{\texttt{C}}, \text{\texttt{G}}\}\} \text{ (shaded in illustration)}
\]

\[
\hat{\epsilon}_m(G) = \{\{\text{\texttt{A}}, \text{\texttt{T}}, \text{\texttt{C}}, \text{\texttt{G}}\}\} \text{ (shaded in illustration)}
\]

\[
\mathbb{V}(G) = 0 \text{ (Definition 2.5 page 28)}
\]

\[
_{\text{\texttt{A}}} \text{ Mendel (1853) (Mendel (1853): gene coding uses discrete symbols), } \text{\texttt{B}} \text{ Watson and Crick (1953a) page 737 (Watson and Crick (1953): gene coding symbols are adenine, thymine, cytosine, and guanine), } \text{\texttt{C}} \text{ Watson and Crick (1953b) page 965, } \text{\texttt{D}} \text{ Pommerville (2013) page 52}
\]

\[
\]
\[ \hat{\mu}(G) \triangleq \arg \min_{x \in G} \max_{y \in G} d(x, y) P(y) \]  
by definition of \( \hat{\mu} \) (Definition 2.3 page 28)

\[ = \arg \min_{x \in G} \max_{y \in G} d(x, y) \frac{1}{4} \]  
by definition of \( G \)

\[ = \arg \min_{x \in G} \max_{y \in G} d(x, y) \frac{1}{4} \]  
because \( f(x) = \frac{1}{4} x \) is strictly isotone and by Lemma 1.7 (page 19)

\[ = \arg \min_{x \in G} \{1, 1, 1, 1\} \]  
because for the discrete metric (Definition D.8 page 155), max \( d = 1 \)

\[ \hat{\mu}_s(G) \triangleq \arg \min_{x \in G} \sum_{y \in G} d(x, y) P(y) \]  
by definition of \( \hat{\mu}_s \) (Definition 2.4 page 28)

\[ = \arg \min_{x \in G} \sum_{y \in G} d(x, y) \frac{1}{6} \]  
by definition of \( G \)

\[ = \arg \min_{x \in G} \sum_{y \in G} d(x, y) \frac{1}{6} \]  
because \( f(x) = \frac{1}{4} x \) is strictly isotone and by Lemma 1.7 (page 19)

\[ = \arg \min_{x \in G} \begin{Bmatrix} 0 + 1 + 1 + 1 \\ 1 + 0 + 1 + 1 \\ 1 + 1 + 0 + 1 \\ 1 + 1 + 1 + 0 \end{Bmatrix} = \arg \min_{x \in G} \begin{Bmatrix} 3 \\ 3 \\ 3 \\ 3 \end{Bmatrix} = \{1, 2, 3, 4\} \]

\[ \text{Var}(G) \triangleq \sum_{x \in G} [d(\hat{\mu}(G), x)]^2 P(x) \]  
by definition of \( \text{Var} \) (Definition 2.5 page 28)

\[ = \sum_{x \in G} (0) \frac{1}{6} \]  
because \( \hat{\mu}(G) = G \)

\[ = 0 \]  
by field property of additive identity element 0

\[ \Rightarrow \]

### 2.2 Random variables on outcome subspaces

#### 2.2.1 Definitions

The traditional random variable (Definition 1.22 page 9) is a mapping from a probability space (Definition 1.21 page 8) to the real line (Definition 1.39 page 22). This paper extends this definition to include functions with additional structure in the domain and expanded structure in the range (next definition).

**Definition 2.13.** A function \( X \in H^G \) (Definition 1.6 page 6) is an outcome random variable if \( G \) is an outcome subspace (Definition 2.1 page 27) and \( H \) is an ordered quasi-metric space (Definition 1.38 page 22).

The definitions of outcome expected value and outcome variance (next definition) of an outcome random variable are, in essence, identical to the outcome center (Definition 2.14 page 46) and outcome variance (Definition 2.14 page 46) of outcome subspaces (Definition 2.1 page 27) that outcome random variables map from and by induction, to.

**Definition 2.14.** Let \( G \) be an outcome subspace (Definition 2.1 page 27), \( H \) an ordered quasi-metric space (Definition 1.38 page 22), and \( X \) be an outcome random variable (Definition 2.13 page 46) in \( H^G \). Let \( H \triangleq (\Omega, \leq, d, P) \) be the outcome subspace induced by \( H, G, \) and \( X \). Let \( \mathbb{E}_X \) be a function from \( \Omega \) to...
the power set \(2^\Omega\).

The **outcome expected value** \( \bar{E}(X) \) of \( X \) is \( \bar{E}(X) \triangleq \arg \min \max_{x \in \Omega} d(x, y) P(y) \).

The **outcome variance** \( \bar{\text{Var}}(X; E_x) \) of \( X \) is \( \bar{\text{Var}}(X) \triangleq \sum_{x \in \Omega} d^2 \left( E_x(X), x \right) P(x) \).

Moreover, \( \bar{\text{Var}}(X) \triangleq \bar{\text{Var}}(X; \bar{E}) \), where \( \bar{E} \) is the outcome expected value function.

### 2.2.2 Properties

**Theorem 2.1.** Let \( X \in H^G \) be a random variable (Definition 2.13 page 46) on an ordered quasi-metric space (Definition 1.38 page 22) \( H \). Let \( H \triangleq (\Omega, \leq, d, \hat{P}) \) be the outcome subspace (Definition 2.1 page 27) induced by \( H \), \( G \), and \( X \). Let \( \text{Var}(X) \) be the traditional variance (Definition 1.23 page 9) of \( X \). Let \( \text{Var}(X) \) be the outcome subspace variance of \( X \) (Definition 2.14 page 46).

\[
\{ H \triangleq (\mathbb{R}, |\cdot|, \leq) \text{ is the real line (Definition 1.39 page 22)} \} \quad \implies \quad \{ \text{Var}(X; E) = \text{Var}(X) \}
\]

**Proof:**

\[
\text{Var}(X; E) \triangleq \sum_{x \in H} d^2(E_x(X), x) P(x) \quad \text{by definition of} \text{Var} \text{ (Definition 2.14 page 46)}
\]

\[
= \sum_{x \in \mathbb{R}} |E(X) - x|^2 P(x) \quad \text{by definition of real line} \ H \text{ (Definition 1.39 page 22)}
\]

\[
= \int_{\mathbb{R}} (x - E(X))^2 P(x) \, dx \quad \text{by definition of Lebesgue integration on} \ \mathbb{R}
\]

\[
= \text{Var}(X) \quad \text{by definition of} \text{Var} \text{ (Definition 1.23 page 9)}
\]

**Remark 2.3.** Despite the correspondence of traditional variance and outcome variance on the real line as demonstrated in Theorem 2.1 (page 47), the situation is different for expected values. Even when both are calculated on the same real line, the traditional expected value \( E(X) \) (Definition 1.23 page 9) and the outcome expected value \( \bar{E}(X) \) (Definition 2.14 page 46) don’t always yield the same value. Demonstrations of this include Example 2.14 (page 54) and Example 2.17 (page 59). However, there is one common situation in which the two statistics do correspond (next theorem).

**Theorem 2.2.** Let \( X, H, G \) be defined as in Theorem 2.1 (page 47). Let \( E(X) \) be the traditional expected value (Definition 1.23 page 9) and \( \bar{E}(X) \) the outcome expected value of \( X \) (Definition 2.14 page 46).

\[
\{ 1. \ H \triangleq (\Omega, d, \leq) \text{ (Real line definition 1.39 page 22)} \text{ and} \ \\
2. \ P(a - x) = P(a + x) \quad \forall x \in \mathbb{R} \text{ (Symmetric about} a) \} \quad \implies \quad \{ \bar{E}(X) = a = E(X) \}
\]

**Proof:**

\[
\bar{E}(X) \triangleq \arg \min \max_{x \in \mathbb{R}} d(x, y) P(y) \quad \text{by definition of} \bar{E} \text{ (Definition 2.14 page 46)}
\]

\[
= \arg \min \max_{x \in \mathbb{R}} |x - y| P(y) \quad \text{by definition of real line (Definition 1.39 page 22)}
\]

\[
= a \quad \text{by Proposition 1.3 (page 9)}
\]

\[
\text{because} \ h(x) \triangleq \max_{y \in \mathbb{R}} |x - y| P(y) \text{ is minimized when} x = a
\]
Theorem 2.3.
Let $G \triangleq (\Omega_G, d_G, \leq_G, P_G)$ be an outcome subspace
(Definition 2.1 page 27).
Let $H \triangleq (\Omega_H, d_H, \leq_H)$ be an ordered quasi-metric space
(Definition 1.38 page 22).
Let $K \triangleq (\Omega_K, d_K, \leq_K)$ be an ordered quasi-metric space
(Definition 1.38 page 22).
Let $X \in H^G$ be a random variable from $G$ onto $H$
(Definition 2.13 page 46).
Let $f \in \Omega_K^H$ be a function from $\Omega_H$ onto $\Omega_K$ (pullback)
(Theorem 1.7 page 21).
Let $\phi \in \mathbb{R}_+$ be a function from $\mathbb{R}$ into $\mathbb{R}$ (pushforward)
(Definition D.11 page 156).
Let $H \triangleq (\Omega_H, d_H, \leq_H, P_H)$ be an outcome subspace induced by $G$, $H$, and $X$.
Let $K \triangleq (\Omega_K, d_K, \leq_K, P_K)$ be an outcome subspace induced by $K$, $H$ and $f$.

\begin{enumerate}
\item $f$ is injective
\item $\phi$ is strictly isotone
\item $d_H(f(x), f(y))P(y) = \phi[d_H(x, y)P(y)]$
\end{enumerate}
\implies \{ \hat{E}[f(X)] = f[\hat{E}(X)] \}

\begin{proof}
\[ \hat{E}[f(X)] = \arg \min_{x \in \Omega_H} \max_{y \in \Omega_K} d_k(x, y)P_K(y) \]
by definition of $\hat{E}$ (Definition 2.14 page 46) and $K$.
\[ = f \left[ \arg \min_{x \in \Omega_H} \max_{y \in \Omega_H} d_H(f(x), f(y))P_K(f(y)) \right] \]
by $f$ bijection hypothesis.
\[ = f \left[ \arg \min_{x \in \Omega_H} \max_{y \in \Omega_H} d_H(f(x), f(y))P_H(y) \right] \]
by $f$ bijection hypothesis.
\[ = f \left[ \arg \min_{x \in \Omega_H} \max_{y \in \Omega_H} \phi[d_H(x, y)P_H(y)] \right] \]
by $d_H$ hypothesis.
\[ = f \left[ \arg \min_{x \in \Omega_H} \max_{y \in \Omega_H} d_H(x, y)P_H(y) \right] \]
by $\phi$ is strictly isotone hypothesis and Lemma 1.7 page 19.
\[ = f \hat{E}(X) \]
by definition of $\hat{E}$ (Definition 2.14 page 46) and $X$.
\end{proof}

Corollary 2.1. Let $H$ be an ordered metric space (Definition 1.38 page 22) and $X \in H^G$ a random variable
(Definition 2.13 page 46) onto $H$. Let $\mathbb{R}_+$ be the real line ordered metric space
(Definition 1.39 page 22).
\[ H = (\mathbb{R}_+, | \cdot |, \leq) \implies \{ \hat{E}(ax) = a \hat{E}(X) \ \forall a \in \mathbb{R}_+ \} \]

\begin{proof}
\begin{enumerate}
\item Proof for $a = 0$ case:
\[ \hat{E}(0 \cdot X) = \arg \min_{x \in \Omega_H} \max_{y \in \Omega_H} d(x, y)P(y) \]
by definition of $\hat{E}$ (Definition 2.14 page 46)
\[ = \max_{y \in \Omega_H} 0P(y) \]
\[ = 0 \]
\[ = 0 \cdot \hat{E}(X) \]
\item Proof for $a > 0$ case:
\[ d(f(x), f(y))P(y) = |ax - ay|P(y) = |a||x - y|P(y) \]
by nondegenerate property of $d$ (Definition D.7 page 153)
\[ \hat{E}(ax) = a \hat{E}(X) \]
because $f(x) = ax$ is strictly isotone on the real line and by Theorem 2.3 (page 48)
\end{enumerate}
\end{proof}
2.2.3 Problem statement

The traditional random variable $X$ (Definition 1.22 page 9) is a function that maps from a stochastic process to the real line (Definition 1.39 page 22). The traditional expectation value $\mathbb{E}(X)$ of $X$ is then often a poor choice of a statistic when the stochastic process that $X$ maps from is a structure other than the real line or some substructure of the real line. There are two fundamental problems:

1. A traditional random variable $X$ maps to the linearly ordered real line. However, $X$ often maps from a random process that is non-linearly ordered (or even unordered Definition 1.24 page 10, Definition 1.25 page 10).

2. A traditional random variable $X$ maps to the real line with a metric geometry (Remark 2.2 page 32) induced by the usual metric (Definition D.9 page 156). But many random processes have a fundamentally different metric geometry, a common one being that induced by the discrete metric (Definition D.8 page 155).

Thus, the order structure of the domain and range of $X$ are often fundamentally dissimilar, leading to statistics, such as $\mathbb{E}(X)$, that are of poor quality with regards to qualitative intuition and quantitative variance (expected error) measurements, and of dubious suitability for tasks such as decision making, prediction, and hypothesis testing.

Remar 2.4. Unlike in traditional statistical processing, it in general not true that $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$. See Example 2.33 (page 76) for a counter example.

Remark 2.5. A possible solution to the traditional random variable order and metric geometry problem is to allow the random variable to map into the complex plane (Example 1.10 page 23) with the usual metric, rather than into the real line only. However, this is a poor solution, as demonstrated in Example 2.21 (page 64).

2.2.4 Examples

Fair die examples

![Figure 2.2: random variable mappings from the fair die to the real line and integer line](image)

*Example 2.10 (fair die mappings to real line and integer line). Let $G$ be the fair die outcome subspace (Example 2.1 page 29). Let $X \in (\mathbb{R}, \|\cdot\|, \preceq)$ be a random variable (Definition 2.13 page 46) mapping from $G$ to the real line (Definition 1.39 page 22), and $Y \in (\mathbb{Z}, \leq, \|\cdot\|)$ be a random variable (Definition 2.13 page 46) mapping from $G$ to the integer line (Definition 1.40 page 23), as illustrated in Figure 2.2 (page 49). Let $E$ be the traditional expected value function (Definition 1.23 page 9), $\vartheta$ the traditional variance function (Definition 1.23 page 9), $\mathbb{E}$ the outcome expected value function (Definition 2.14 page 46), and $\vartheta$ the outcome variance function.
This yields the following statistics:

**geometry of \( G \):**

\[ \hat{\mathcal{G}}(G) = \{ \square, \lozenge, \ino, \inout, \bigcirc, \bigtriangleup \} \]

**traditional statistics on real line:**

\[ E(X) = 3.5 \quad \text{Var}(X; E) = \frac{35}{12} \approx 2.917 \]

**outcome subspace statistics on real line:**

\[ \hat{E}(X) = \{ 3.5 \} \quad \text{Var}(X; \hat{E}) = \frac{35}{12} \approx 2.917 \]

**outcome subspace statistics on integer line:**

\[ \hat{E}(Y) = \{ 3, 4 \} \quad \text{Var}(Y; \hat{E}) = \frac{5}{6} \approx 1.667 \]

**Proof:**

\[ \hat{\mathcal{G}}(G) = \{ \square, \lozenge, \ino, \inout, \bigcirc, \bigtriangleup \} \quad \text{by Example 2.1 page 29} \]

\[ E(X) \triangleq \sum_{x \in \mathbb{R}} x P(x) \quad \text{by definition of } E \quad \text{(Definition 1.23 page 9)} \]

\[ = \sum_{x \in \mathbb{R}} x \cdot \frac{1}{6} = \frac{1 + 2 + 3 + 4 + 5 + 6}{6} = \frac{21}{6} = \frac{7}{2} = 3.5 \]

\[ \text{Var}(X; E) = \text{Var}(X) \quad \text{by Theorem 2.1 page 47} \]

\[ \triangleq \sum_{x \in \mathbb{R}} [x - E(X)]^2 P(x) \quad \text{by definition of } \text{Var} \quad \text{(Definition 1.23 page 9)} \]

\[ = \sum_{x \in \mathbb{R}} \left( x - \frac{7}{2} \right)^2 \cdot \frac{1}{6} \quad \text{by } E(X) \text{ result} \]

\[ = \left[ \left( 1 - \frac{7}{2} \right)^2 + \left( 2 - \frac{7}{2} \right)^2 + \left( 3 - \frac{7}{2} \right)^2 + \left( 4 - \frac{7}{2} \right)^2 + \left( 5 - \frac{7}{2} \right)^2 + \left( 6 - \frac{7}{2} \right)^2 \right] \cdot \frac{1}{6} \]

\[ = \left[ \frac{1}{22} \times 2 \times 6 \right] \quad \text{by } E(X) \text{ result} \]

\[ = \frac{25 + 9 + 1 + 1 + 9 + 25}{24} = \frac{70}{24} = \frac{35}{12} \approx 2.917 \]

\[ \hat{E}(X) = E(X) \quad \text{by Theorem 2.2 (page 47)} \]

\[ = \left\{ \frac{7}{2} \right\} = \{ 3.5 \} \quad \text{by } E(X) \text{ result} \]

\[ \text{Var}(X; \hat{E}) \triangleq \sum_{x \in \mathbb{R}} d^2(\hat{E}(X), x) P(x) \quad \text{by definition of } \text{Var} \quad \text{(Definition 2.14 page 46)} \]

\[ = \sum_{x \in \mathbb{R}} d^2(E(X), x) P(x) \quad \text{by } E(X) \text{ and } \hat{E}(X) \text{ results} \]

\[ \text{Var}(X; E) \quad \text{by definition of } \text{Var} \quad \text{(Definition 2.14 page 46)} \]

\[ = \frac{35}{12} \approx 2.917 \]

\[ \hat{E}(Y) \triangleq \arg \min_{x \in \mathbb{Z}} \max_{y \in \mathcal{G}} d(x, y) P(y) \quad \text{by definition of } \hat{E} \quad \text{(Definition 2.14 page 46)} \]

\[ = \arg \min_{x \in \mathbb{Z}} \max_{y \in \mathcal{G}} |x - y| \cdot \frac{1}{6} \quad \text{by definition of } \text{integer line} \quad \text{(Definition 1.40 page 23)} \]

\[ = \arg \min_{x \in \mathbb{Z}} \max_{y \in \mathcal{G}} |x - y| \quad \text{by } \phi(x) = \frac{1}{6} x \text{ is strictly isotone} \quad \text{and by Lemma 1.7 page 19} \]

\[ = \frac{1}{2} \sum_{x \in \mathbb{Z}} \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 5 \\ 1 & 0 & 1 & 2 & 3 & 4 \\ 2 & 1 & 0 & 1 & 2 & 3 \\ 3 & 2 & 1 & 0 & 1 & 2 \\ 4 & 3 & 2 & 1 & 0 & 1 \\ 5 & 4 & 3 & 2 & 1 & 0 \end{bmatrix} = \frac{1}{2} \sum_{x \in \mathbb{Z}} \begin{bmatrix} 5 & 4 \\ 3 & 3 \\ 4 & 5 \end{bmatrix} = \frac{1}{2} \sum_{x \in \mathbb{Z}} \begin{bmatrix} 3 \\ 4 \end{bmatrix} \]

\[ = \frac{1}{6} \left( |3 - 1|^2 + |3 - 2|^2 + |3 - 3|^2 + |4 - 4|^2 + |4 - 5|^2 + |4 - 6|^2 \right) \quad \text{by } \hat{E}(Y) \text{ result and definition of } \mathcal{G} \]
\[
-
\frac{1}{6}(4 + 1 + 0 + 0 + 1 + 4) = \frac{10}{6} = \frac{5}{3}
\]

The random variable mappings in Example 2.10 (page 49) have two fundamental problems:

1. The order structure of the **fair die** and the order structure of **real line** are inherently dissimilar in that while the bijective (Definition 1.14 page 7) mapping \( X \) is trivially order preserving (Definition 1.36 page 16), its inverse is not order preserving. And this is a problem. In the linearly ordered (Definition 1.25 page 10) range of \( X \), it is true that \( X(\square) = 1 < 2 = X(\varnothing) \). But in the unordered domain of \( X \ \{ \square, \varnothing, \ldots, \exists \} \), it is not true that \( \square < \varnothing \); rather \( \square \) and \( \varnothing \) are simply symbols without order. This causes problems when we attempt to use the random variable to make statistical inferences involving moments (Definition 2.2 page 27). The traditional expected value (Definition 1.23 page 9) of a fair die (Example 2.1 page 29) is \( E(X) = \frac{1}{6}(1 + 2 + \cdots + 6) = 3.5 \). This implies that we expect the outcome of \( \square \) or \( \varnothing \) more than we expect the outcome of say \( \square \) or \( \varnothing \). But these results have no relationship with reality or with intuition because the values of a fair die are merely symbols. For a fair die, we would not expect any pair of values equally. We would not expect the outcome \( [\square \) or \( \varnothing \) more than we would expect the outcome \( [\square \) or \( \varnothing \) \], or more than we would expect any other outcome pair.

2. The metric geometry (Remark 2.2 page 32) of the **fair die outcome subspace** is very dissimilar to the metric geometry of the real line (Definition 1.39 page 22) that it is mapped to by the random variable \( X \). And this is a problem. In the metric geometry of the fair die induced by the discrete metric (Definition D.8 page 155), \( \square \) is no closer to \( \varnothing \) than it is to \( \diamond \) \( d(\square, \varnothing) = 1 = d(\square, \diamond) \)). However in the metric geometry of the real line induced by the usual metric \( d(x, y) \triangleq |x - y| \) (Definition D.9 page 156), \( X(\square) = 1 \) is closer to \( X(\varnothing) = 2 \) than it is to \( X(\diamond) = 3 \) \( |1 - 2| = 1 \neq 2 = |1 - 3| \).

\[\text{(A) mapping to isomorphic structure} \quad \text{(B) mapping to extended structure}\]

**Figure 2.3:** order preserving random variable mappings from fair die

**Example 2.11** (fair die mapping to isomorphic structure). Let \( G \triangleq (\{ \square, \varnothing, \diamond, \centerdot, \bullet, \exists \}, d, \varnothing, P) \) be a fair die outcome subspace (Example 2.1 page 29), and \( H \triangleq (\{ 1, 2, 3, 4, 5, 6 \}, d, \varnothing) \) be an unordered metric space (Definition 1.38 page 22). Example 2.10 (page 49) presented mappings from \( G \) to structures with structures dissimilar to \( G \). Figure 2.3 page 51 (A) illustrates a mapping to the isomorphic structure \( H \triangleq (\{ 1, 2, 3, 4, 5, 6 \}, d, \varnothing, X(P)) \), yielding the following statistics:

\[
\bar{E}(X) = \{1, 2, 3, 4, 5, 6\} \quad \text{Var}(X) = 0
\]

Here, \( \bar{E}(X) \) equals the entire base set of \( H \), indicating a statistic carrying no information about an expected outcome. That is, there is no best guess concerning outcome. This is much different than the traditional probability of 3.5 (Example 2.10 page 49) which deceptively suggests a likely outcome of \( \square \) or \( \varnothing \). And one could easily argue that no information is much better than misleading information.
\[ E(X) = \arg \min_{x \in H} \max_{y \in H} d(x, y) P(y) \]

by definition of \( E \) (Definition 2.14 page 46)

\[ = \arg \min_{x \in H} \max_{y \in H} d(x, y) P(y) \quad \text{because } G \text{ and } H \text{ are isomorphic} \]

\[ = X[\mathcal{L}(G)] \quad \text{by definition of } \mathcal{L} \text{ (Definition 2.23 page 28)} \]

\[ = X[\{\square, \square, \square, \square, \square, \square\}] \quad \text{by Example 2.10 (page 49)} \]

\[ = \{1, 2, 3, 4, 5, 6\} \quad \text{by definition of } X \]

\[ \mathbb{V} \text{ar}(X) = \sum_{x \in H} \left( E(X), x \right) P(x) \]

by definition of \( \mathbb{V} \text{ar} \) (Definition 2.14 page 46)

\[ = \sum_{x \in G} d^2(X(G), x) P(x) \quad \text{because } G \text{ and } H \text{ are isomorphic} \]

\[ \triangleq \mathbb{V} \text{ar}(G) \quad \text{by definition of } \mathbb{V} \text{ar} \text{ (Definition 2.5 page 28)} \]

\[ = 0 \quad \text{by Example 2.10 (page 49)} \]

Although all the coefficients of the polynomial equation \( x^2 - 2x + 2 = 0 \) are in the set of real numbers \( \mathbb{R} \), the solutions of the equation \( x = 1 + i \) and \( x = 1 - i \) are not. Rather, the two solutions are in the complex plane \( \mathbb{C} \) (Example 1.10 page 23), of which \( \mathbb{R} \) is a substructure. This is an example of extending a structure (from \( \mathbb{R} \) to \( \mathbb{C} \) ) to achieve more useful results. The same idea can be applied to a random variable \( X \in H^G \). The definition of an outcome random variable (Definition 2.13 page 46) does not require a bijection between \( G \) and \( H \); rather, it only requires that the mapping be “into” the base set of \( H \) (Definition 1.14 page 7). In Example 2.11 (page 51) in which \( G \) is isomorphic to \( H \), the expected value of \( X \) is a set with six values. However, we could extend \( H \), while still preserving the order and metric geometry of \( G \), to produce a random variable with a simpler expected value (next example).

**Example 2.12** (fair die mapping with extended range). Let \( G \triangleq \{\square, \square, \square, \square, \square, \square\} \), \( d, \varnothing, P \) be a fair die outcome subspace (Example 2.1 page 29), and \( H \triangleq \{1, 2, 3, 4, 5, 6, 0\}, p, \varnothing \) be an unordered metric space (Definition 1.38 page 22). Figure 2.3 page 51 (B) illustrates a random variable mapping \( X \) from \( G \) to the extended structure \( H \), yielding the following statistics:

\[ E(X) = \{0\} \quad \text{Var}(X) = \frac{1}{4} \]

As in Example 2.11, order and metric geometry are still preserved. Here, an expected value of \( \{0\} \) simply means that no real physical value is expected more or less than any other real physical value. Note also that the variance (expected error) is more than 11 times smaller than that of the corresponding statistical estimates on the real line \( \|t_2 \text{ versus } 30\|_2 \) Example 2.10 page 49).

\[ E(X) = \arg \min_{x \in H} \max_{y \in H} d(x, y) P(y) \quad \text{by definition of } E \text{ (Definition 2.14 page 46)} \]

\[ = \arg \min_{x \in H} \max_{y \in H} d(x, y) P(y) \quad \text{because } P(0) = 0 \]

\[ = \arg \min_{x \in H} \max_{y \in H} d(x, y) \frac{1}{6} \quad \text{because } f(x) = \frac{1}{6} x \text{ is strictly isotone} \text{ and by Lemma 1.7 (page 19)} \]

\[ = \arg \min_{x \in H} \max_{y \in H} \begin{bmatrix} d(1, 1) & d(1, 2) & \cdots & d(1, 6) \\ d(2, 1) & d(2, 2) & \cdots & d(2, 6) \\ d(3, 1) & d(3, 2) & \cdots & d(3, 6) \\ \vdots & \vdots & \ddots & \vdots \\ d(5, 1) & d(5, 2) & \cdots & d(5, 6) \\ d(6, 1) & d(6, 2) & \cdots & d(6, 6) \\ d(0, 1) & d(0, 2) & \cdots & d(0, 6) \end{bmatrix} = \arg \min_{x \in H} \max_{y \in H} \begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 2 \\ 1 & 0 & 1 & 1 & 2 & 1 \\ 1 & 1 & 0 & 2 & 1 & 1 \\ 1 & 2 & 1 & 1 & 0 & 1 \\ 2 & 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix} = \arg \min_{x \in H} \begin{bmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \end{bmatrix} \]
CHAPTER 2. STOCHASTIC PROCESSING ON WEIGHTED GRAPHS

\[
\text{Var}(X) = \sum_{x \in H} d^2(E(X), x) P(x)
\]

by definition of \text{Var} (Definition 2.14 page 46)

\[
= \sum_{x \in H} d^2(\{0\}, x) P(x)
\]

by \(\hat{E}(X)\) result

\[
= \sum_{x \in H \setminus \{0\}} d^2(\{0\}, x) \frac{1}{6}
\]

by definition of \(G\)

\[
= 6 \left( \frac{1}{2} \right)^2 \cdot \frac{1}{6} = \frac{1}{4}
\]

by definition of \(H\)

\[\Rightarrow\]

Real die examples

![Diagram showing random variable mappings from the real die outcome subspace to several ordered metric spaces](Example 2.13 page 53)

**Example 2.13** (real die mappings). Let \(G\) be the real die outcome subspace (Example 2.2 page 31). Let \(W, X, Y\) and \(Z\) be random variable (Definition 2.13 page 46) mappings as illustrated in Figure 2.4 (page 53). Let \(E, \text{Var}, \hat{E}, \text{Var}\) be defined as in Example 2.10 (page 49). This yields the following statistics:

- **geometry of \(G\):** \(\hat{E}(G) = \{O, R, \hat{O}, \hat{R}, 0, 1\}\)
- **traditional statistics on real line:** \(E(W) = 3.5\) \(\text{Var}(W; E) = \frac{35}{12} \approx 2.917\)
- **outcome subspace statistics on real line:** \(\hat{E}(W) = \{3.5\}\) \(\text{Var}(W; \hat{E}) = \frac{35}{12} \approx 2.917\)
- **outcome subspace statistics on integer line:** \(E(X) = \{3, 4\}\) \(\text{Var}(X; E) = \frac{5}{12} \approx 1.667\)
- **outcome subspace statistics on isomorphic structure:** \(\hat{E}(Y) = \{1, 2, \ldots, 6\}\) \(\text{Var}(Y; \hat{E}) = 0\)
- **outcome subspace statistics on extended structure:** \(\hat{E}(Z) = \{0\}\) \(\text{Var}(Z; \hat{E}) = 1\)

Similar to Example 2.11 (page 51), the statistic \(\hat{E}(Z) = \{0\}\) indicates a statistic carrying no information about an expected outcome. Again, one could easily argue that no information is much better than misleading information.
PROOF:

\[ \hat{\mathbb{E}}(G) = \{ \Box, \lozenge, \Diamond, \bigstar, \bigcirc \} \]  
by Example 2.2 (page 31)

\[ E(W) = \frac{7}{2} = 3.5 \]  
by E(X) result of Example 2.10 page 49

\[ \hat{\mathbb{V}}ar(W; E) = \frac{35}{12} \approx 2.917 \]  
by \[ \mathbb{V}ar(X) \] result of Example 2.10 page 49

\[ E(W) = \{ 3.5 \} \]  
by \[ \mathbb{E}(X) \] result of Example 2.10 page 49

\[ \hat{\mathbb{V}}ar(W; \hat{\mathbb{E}}) = \frac{35}{12} \approx 2.917 \]  
by \[ \mathbb{V}ar(X; \hat{\mathbb{E}}) \] result of Example 2.10 page 49

\[ \hat{\mathbb{E}}(X) = \{ 3, 4 \} \]  
by \[ \mathbb{E}(Y) \] result of Example 2.10 page 49

\[ \hat{\mathbb{V}}ar(X; \hat{\mathbb{E}}) = \frac{5}{3} \approx 1.667 \]  
by \[ \mathbb{V}ar(Y; \hat{\mathbb{E}}) \] result of Example 2.10 page 49

\[ \hat{\mathbb{E}}(Y) = \{ 1, 2, 3, 4, 5, 6 \} \]  
by \[ \hat{\mathbb{E}}(G) \] result of Example 2.2 page 31

\[ \hat{\mathbb{V}}ar(Y; \hat{\mathbb{E}}) \triangleq \sum_{x \in \mathbb{H}} d^2(\hat{\mathbb{E}}(Y), x) P(x) \]  
by definition of \[ \hat{\mathbb{V}}ar \] (Definition 2.14 page 46)

\[ = \sum_{x \in \mathbb{H}} d^2(H, x) P(x) \]  
by \[ \hat{\mathbb{E}}(Y) \] result

\[ = \sum_{x \in \mathbb{H}} 0^2 x P(x) \]  
by nondegenerate property of quasi-metrics (Definition D.6 page 153)

\[ = 0 \]

\[ \hat{\mathbb{E}}(Z) \triangleq \arg \min \max_{x \in \mathbb{K}} d(x, y) P(y) \]  
by definition of \[ \hat{\mathbb{E}} \] (Definition 2.14 page 46)

\[ = \arg \min_{x \in \mathbb{K}} \max_{y \in \mathbb{K} \setminus \{0\}} d(x, y) P(y) \]  
because \[ P(0) = 0 \]

\[ = \arg \min_{x \in \mathbb{K}} \max_{y \in \mathbb{K} \setminus \{0\}} \frac{1}{6} d(x, y) \]  
by definition of \[ \hat{\mathbb{E}} \] and \[ \hat{\mathbb{E}} \]

\[ = \arg \min_{x \in \mathbb{K}} \max_{y \in \mathbb{K} \setminus \{0\}} \begin{pmatrix} d(1, 1) \cdots d(1, 6) \\ d(2, 1) \cdots d(2, 6) \\ \vdots \end{pmatrix} \]  
\[ = \begin{pmatrix} 0 & 1 & 1 & 1 & 2 \\ 0 & 1 & 1 & 2 & 1 \\ 1 & 1 & 0 & 2 & 1 \\ 1 & 1 & 2 & 0 & 1 \\ 1 & 2 & 1 & 1 & 0 \\ 2 & 1 & 1 & 1 & 1 \end{pmatrix} \]  
\[ = \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \end{pmatrix} \]  
\[ = \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \end{pmatrix} \]

\[ \hat{\mathbb{V}}ar(Z) \triangleq \sum_{x \in \mathbb{K}} d^2(\hat{\mathbb{E}}(Z), x) P(x) \]  
by definition of \[ \hat{\mathbb{V}}ar \] (Definition 2.14 page 46)

\[ = \sum_{x \in \mathbb{K} \setminus \{0\}} d^2(\{0\}, x) P(x) \]  
by \[ \hat{\mathbb{E}}(Z) \] result

\[ = 6 \times 1^2 \times \frac{1}{6} = 1 \]  
by \[ \hat{\mathbb{E}}(Z) \] result

Example 2.14 (weighted die mappings). Let \( G \) be weighted die outcome subspaces (Example 2.3 page 33), and \( X \) and \( Y \) be random variables, as illustrated in Figure 2.5 (page 55). Let \( \mathbb{E}, \mathbb{V}ar, \hat{\mathbb{E}}, \) and \( \hat{\mathbb{V}}ar \) be defined as in Example 2.10 (page 49). This yields the following statistics:

\[ \hat{\mathbb{E}}(G) = \{ \Box \} \]

traditional statistics on real line:

\[ \mathbb{E}(X) = 3 \]

\[ \mathbb{V}ar(X) \approx 3.57 \]

outcome subspace statistics on real line:

\[ \hat{\mathbb{E}}(X) = \{ \Box \} \approx \{ 3.57 \} \]

\[ \hat{\mathbb{V}}ar(X; \hat{\mathbb{E}}) \approx 1.43 \]

outcome subspace stats. on isomorphic structure

\[ \hat{\mathbb{E}}(Y) = \{ 4 \} \]

\[ \hat{\mathbb{V}}ar(X; \hat{\mathbb{E}}) \approx 0.337 \]

The statistic \( \mathbb{E}(X) = 3 \) evaluated on the real line is arguably very poor because it suggests that we "expect" the event \( \Box \) rather than \( \Box \), even though \( P(\Box) \) is very large, \( P(\Box) \) is very small, and the
physical distance \(d(\bullet, \bigcirc) = 2\) on the die from \(\bullet\) to \(\bigcirc\) is twice as much as it is to any of the other four die faces. If we retain use of the real line but replace the traditional expected value \(E(X)\) with the outcome expected value \(\bar{E}(X)\), a small but significant improvement is made \((\bar{\text{Var}}(X; \bar{E}) \approx 1.143 < 1.43 = \text{Var}(X; E))\). Arguably a better choice still is to abandon the real line altogether in favor of the isomorphic structure \(K\) and the statistic \(\bar{E}(Y) = \{4\}\) evaluated on \(K\), yielding not only an intuitively better result but also a variance \(\bar{\text{Var}}(Y; \bar{E})\) that is more than 4 times smaller than that of \(E(X)\) \((\bar{\text{Var}}(Y; \bar{E}) \approx 0.337 < 1.43 = \bar{\text{Var}}(X; E))\).

\[\begin{align*}
\text{Proof:} & \\
\hat{E}(G) &= \{ \bullet \} & \text{by Example 2.3 (page 33)} \\
E(X) &= \int_{\mathbb{R}} xP(x) \, dx & \text{by definition of } E \text{ (Definition 1.23 page 9)} \\
&= \sum_{x \in \mathbb{Z}} xP(x) & \text{by definition of } P \\
&= 1 \times \frac{1}{10} + 2 \times \frac{1}{20} + 3 \times \frac{1}{30} + 4 \times \frac{3}{5} + 5 \times \frac{1}{50} + 6 \times \frac{1}{30} \\
&= \frac{1}{300} (1 \times 30 + 2 \times 15 + 3 \times 10 + 4 \times 180 + 5 \times 6 + 6 \times 10) = \frac{900}{300} = 3 \\
\text{Var}(X; E) &= \text{Var}(X) & \text{by Theorem 2.1 page 47} \\
&= \int_{\mathbb{R}} [x - E(X)]^2P(x) & \text{by definition of } \text{Var} \text{ (Definition 1.23 page 9)} \\
&= \sum_{x \in \mathbb{Z}} [x - E(X)]^2P(x) & \text{by definition of } P \\
&= \frac{1}{10} (1 - 3)^2 + \frac{1}{20} (2 - 3)^2 + \frac{1}{30} (3 - 3)^2 + \frac{3}{5} (4 - 3)^2 + \frac{1}{50} (5 - 3)^2 + \frac{1}{30} (6 - 3)^2 \\
&= \frac{1}{300} (120 + 15 + 0 + 180 + 24 + 90) = \frac{429}{300} = \frac{143}{100} = 1.43 \\
\bar{E}(X) &= \arg\min_{x \in \mathbb{R}} \max_{y \in \mathbb{R}} d(x, y)P(y) & \text{by definition of } \bar{E} \text{ (Definition 2.14 page 46)} \\
&= \arg\min_{x \in \mathbb{R}} \max_{y \in \mathbb{R}} |x - y|P(y) & \text{by definition standard metric on real line (Definition 1.39 page 22)} \\
&= \arg\min_{x \in \mathbb{R}} \left\{ \frac{|x - 1|}{10} \right\} & \text{for } \frac{25}{7} \leq x \leq \frac{23}{3} \\
&= \left\{ \frac{25}{7} \right\} \approx \{3.5714\} & \text{otherwise} \\
\bar{\text{Var}}(X; \bar{E}) &= \sum_{x \in \mathbb{R}} d^2(\bar{E}(X), x)P(x) & \text{by definition of } \bar{\text{Var}} \text{ (Definition 2.14 page 46)} \\
&= \sum_{x \in \mathbb{R}} d^2 \left( \frac{25}{7}, x \right)P(x) & \text{by } \bar{E}(X) \text{ result}
\end{align*}\]
\[
\begin{align*}
\bar{X} &= (\frac{25}{7} - 1)^2 \cdot \frac{1}{10} + (\frac{25}{7} - 2)^2 \cdot \frac{1}{20} + (\frac{25}{7} - 3)^2 \cdot \frac{1}{30} + (\frac{25}{7} - 4)^2 \cdot \frac{3}{5} + (\frac{25}{7} - 5)^2 \cdot \frac{1}{50} + (\frac{25}{7} - 6)^2 \cdot \frac{1}{30} \\
&= \frac{16805}{49 \cdot 300} = \frac{3361}{2940} = 1.143 \\
E(Y) &= Y(\hat{U}(G)) \\
&= Y(\{5\}) \\
&= \{4\} \\
\text{Var}(Y; \hat{E}) &= \text{Var}(G) = \frac{101}{300} \approx 0.337
\end{align*}
\]

because \( G \) and \( H \) are isomorphic

by Example 2.3 (page 33)

by definition of \( Y \)

by Example 2.3 (page 33)

**Spinner examples**

Figure 2.6: Six value fair spinner with assorted random variable mappings (Example 2.15 page 56)

**Example 2.15 (spinner mappings).** A six value board game spinner has a cyclic structure as illustrated in Figure 2.6 (page 56). Again, the order and metric geometry of the real line mapped to by the random variable \( X \) is very dissimilar to that of the outcome subspace that it is supposed to represent. Therefore, statistical inferences based on \( X \) will likely result in values that are arguably unacceptable. Both random variables \( Y \) and \( Z \) map to structures in which order and metric geometry are preserved. The mappings yield the following statistics:

- **geometry of \( G \):**
  \[ \hat{U}(G) = \{1, 2, 3, 4, 5, 6\} \]

- **traditional statistics on real line:**
  \[ E(W) = 3.5 \quad \text{Var}(W; \hat{E}) = \frac{35}{12} \approx 2.917 \]

- **outcome subspace statistics on real line:**
  \[ \hat{E}(W) = \{3.5\} \quad \text{Var}(W; \hat{E}) = \frac{35}{12} \approx 2.917 \]

- **outcome subspace statistics on integer line:**
  \[ \hat{E}(X) = \{3, 4\} \quad \text{Var}(X; \hat{E}) = \frac{12}{20} = 0.667 \]

- **outcome subspace stats. on isomorphic structure:**
  \[ \hat{E}(Y) = \{1, 2, 3, 4, 5, 6\} \quad \text{Var}(Y; \hat{E}) = 0 \]

- **outcome subspace stats. on extended structure:**
  \[ \hat{E}(Z) = \{0\} \quad \text{Var}(Z; \hat{E}) = \frac{1}{2} = 2.25 \]

**Proof:**

\[ \hat{U}(G) = \{1, 2, 3, 4, 5, 6\} \]

by Example 2.4 (page 35)

\[ E(W) = \sum_{x \in \mathbb{R}} xP(x) \]

by definition of \( E \) (Definition 1.23 page 9)

\[ = \frac{2}{7} = 3.5 \]

by fair die example Example 2.10 (page 49)

\[ \text{Var}(W; \hat{E}) = \text{Var}(X) \]

by Theorem 2.1 page 47
\[ \sum_{x \in \mathbb{R}} (x - E(X))^2 P(x) \] 
by definition of \( \text{Var} \) (Definition 2.13 page 9)

\[ = \frac{35}{12} \approx 2.917 \]
by \textit{fair die} example Example 2.10 (page 49)

\( \bar{E}(W) = E(W) \)

\[ = \{3.5\} \]
because on \textit{real line}, \( P \) is \textit{symmetric}, and by Theorem 2.2 page 47 by \( E(W) \) result

\[ \text{Var}(W; \bar{E}) = \text{Var}(W; E) \]
because \( \bar{E}(W) = E(W) \)

\[ = \frac{35}{12} \approx 2.917 \]
by \( \text{Var}(W; E) \) result

\[ \bar{E}(X) \triangleq \arg \min_{x \in \mathbb{Z}} \max_{y \in \mathbb{Z}} d(x, y) P(y) \] 
by definition of \( \bar{E} \) (Definition 2.14 page 46)

\[ \bar{E}(X) = \arg \min_{x \in \mathbb{Z}} \max_{y \in \mathbb{Z}} \left\{ \begin{array}{c} d(1, 1) \ldots d(1, 6) \\ d(2, 1) \ldots d(2, 6) \\ d(3, 1) \ldots d(3, 6) \\ d(4, 1) \ldots d(4, 6) \\ d(5, 1) \ldots d(5, 6) \\ d(6, 1) \ldots d(6, 6) \\ d(0, 1) \ldots d(0, 6) \end{array} \right\} = \arg \min_{x \in \mathbb{Z}} \max_{y \in \mathbb{Z}} \left\{ \begin{array}{c} 0 \ 1 \ 2 \ 3 \ 2 \ 1 \\ 1 \ 0 \ 1 \ 2 \ 3 \ 2 \\ 2 \ 1 \ 0 \ 1 \ 2 \ 3 \\ 3 \ 2 \ 1 \ 0 \ 1 \ 2 \\
\right\} = \left\{ \begin{array}{c} 3 \ 3 \ 3 \ 3 \ 3 \ 3 \\ 2 \ 2 \ 2 \ 2 \ 2 \ 2 \\ \end{array} \right\} \]

\[ \text{Var}(Z) \triangleq \sum_{x \in \mathbb{Z}} d^2(\hat{\mathcal{G}}, x) P(x) \] 
by definition of \( \text{Var} \) (Definition 2.14 page 46)

\[ = \sum_{x \in \mathbb{Z}} d^2(0, x) P(x) \]
by \( \bar{E}(X) \) result

\[ = \sum_{x \in \mathbb{Z}} \left( \frac{3}{2} \right)^2 \frac{1}{6} = |H \setminus \{0\}| \left( \frac{3}{2} \right)^2 \frac{1}{6} = 6 \left( \frac{3}{2} \right)^2 \frac{1}{6} = \frac{9}{4} \]

\[ \Rightarrow \]

\textbf{Example 2.16 (weighted spinner mappings).} Let \( G \) be \textit{weighted spinner outcome subspace} (Example 2.5 page 37) with random variable mappings as illustrated in Figure 2.7 (page 58). This yields the following statistics:

- geometry of \( G \):
  \[ \hat{\mathcal{G}}(G) = \{ 0, \hat{G} \} \]

- traditional statistics on real line \((\mathbb{R}, |\cdot|, \leq)\):
  \[ E(X) = \{ 3.5 \} \text{ Var}(W; E) = \{ 4.25 \} \]

- outcome subspace statistics on real line \((\mathbb{R}, |\cdot|, \leq)\):
  \[ \bar{E}(X) = \{ 3.5 \} \text{ Var}(W; \bar{E}) = \{ 1.67 \} \]

- outcome subspace statistics on isomorphic structure \( H \):
  \[ \bar{E}(Y) = \{ 1.6 \} \text{ Var}(Y; \bar{E}) = \{ 1.85 \} \]

Note that based on the variance values, the statistic \( \bar{E}(Z) \) on the continuous ring \( K \) is arguably a much better statistic than \( \bar{E}(X) \) on the (continuous) real line \((\mathbb{R}, |\cdot|, \leq)\).
\[ 
\mathcal{G} = \{1, 6\} \]  
by \textit{weighted spinner outcome subspace example} (Example 2.5 page 37)

\[ 
E(X) = \sum_{x \in Z} xP(x) 
= 1 \times \frac{3}{10} + 2 \times \frac{1}{10} + 3 \times \frac{1}{10} + 4 \times \frac{1}{10} + 5 \times \frac{1}{10} + 6 \times \frac{1}{10} 
= \frac{1}{10}(1 \times 3 + 2 \times 3 + 4 \times 5 + 6 \times 3) = \frac{35}{10} = \frac{7}{2} = 3.5 
\]

\[ 
\text{Var}(X; E) = \text{Var}(X) 
\triangleq \sum_{x \in Z} [x - E(X)]^2P(x) 
\]  
by definition of \text{Var} (Definition 1.23 page 9)

\[ 
\begin{align*} 
&= \left(1 - \frac{7}{2}\right)^2 \frac{3}{10} + \left(2 - \frac{7}{2}\right)^2 \frac{1}{10} + \left(3 - \frac{7}{2}\right)^2 \frac{1}{10} + \left(4 - \frac{7}{2}\right)^2 \frac{1}{10} + \left(5 - \frac{7}{2}\right)^2 \frac{1}{10} + \left(6 - \frac{7}{2}\right)^2 \frac{3}{10} \\
&= \frac{1}{10} \left(\left(-\frac{5}{2}\right)^2 \times 3 + \left(-\frac{3}{2}\right)^2 + \left(-\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{3}{2}\right)^2 + \left(\frac{5}{2}\right)^2 \times 3\right) \\
&= \frac{1}{40} \left(75 + 9 + 1 + 9 + 75\right) = \frac{170}{40} = \frac{17}{4} = 4.25 
\end{align*} 
\]

\[ 
\mathbb{E}(X) = E(X) 
\]  
because on \textit{real line}, \(P\) is \textit{symmetric}, and by Theorem 2.2 page 47

\[ 
= 3.5 
\]  
by \(E(X)\) result

\[ 
\text{Var}(X; \mathbb{E}) = \text{Var}(X; E) 
\]  
by \(E(X)\) and \(\mathbb{E}(X)\) results

\[ 
= \frac{17}{4} = 4.25 
\]  
by \(\text{Var}(X; E)\) result

\[ 
\mathbb{E}(Y) = Y[\mathbb{E}(\mathcal{G})] 
\]  
because \(\mathcal{G}\) and \(H\) are \textit{isomorphic} under \(Y\)

\[ 
= Y[\mathbb{E}(\{1, 6\})] 
\]  
by \(\mathbb{E}(\mathcal{G})\) result

\[ 
= \mathbb{E}(\{1, 6\}) 
\]  
by definition of \(Y\)

\[ 
\text{Var}(Y; \mathbb{E}) = \text{Var}(\mathcal{G}) 
\]  
because \(\mathcal{G}\) and \(H\) are \textit{isomorphic} under \(Y\)

\[ 
= \frac{5}{3} \approx 1.667 
\]  
by \textit{weighted spinner outcome subspace example} (Example 2.5 page 37)

\[ 
\mathbb{E}(Z) \triangleq \arg \min_{x \in K} \max_{y \in K} d(x, y) P(y) 
\]  
by definition of \(\mathbb{E}\) (Definition 2.14 page 46)

\[ 
= 0.5 
\]

\[ 
\text{Var}(Z; \mathbb{E}) \triangleq \sum_{x \in K} d^2(\mathbb{E}(Z), x) P(x) 
\]  
by definition of \(\text{Var}\) (Definition 2.14 page 46)

\[ 
= \sum_{x \in K} d^2\left(\frac{3}{2}, x\right) P(x) 
\]  
by \(\mathbb{E}(Z)\) result

\[ 
= \left(\frac{1}{2}\right)^2 \left(\frac{3}{2}\right)^2 \frac{1}{10} + \left(\frac{3}{2}\right)^2 \left(\frac{1}{2}\right)^2 + \left(\frac{5}{2}\right)^2 \left(\frac{5}{2}\right)^2 \frac{1}{10} + \left(\frac{5}{2}\right)^2 \left(\frac{1}{2}\right)^2 + \left(\frac{3}{2}\right)^2 \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 \frac{3}{10} 
\]
Pseudo-random number generator (PRNG) examples

Figure 2.8: LCG mappings to linear (X), non-linear discrete (Y) and non-linear continuous (Z) ordered metric spaces (Example 2.17 page 59)

Example 2.17 (LCG mappings, standard ordering). The equation \( x_{n+1} = (7x_n + 5) \mod 9 \) with \( x_0 = 1 \) is a linear congruential (LCG) pseudo-random number generator (PRNG) that has full period\(^5\) of 9 values. These 9 values can be mapped, using a surjective (Definition 1.14 page 7) function \( s \in G^G \) to the 5 element set \{0, 1, 2, 3, 4\} to “shape” the distribution from a uniform distribution to non-uniform:\(^6\)

\[
\begin{array}{cccccccccccc}
 n & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & \cdots \\
x_n & 1 & 5 & 8 & 7 & 0 & 5 & 6 & 2 & 1 & 3 & 8 & \cdots \\
y_n & 1 & 3 & 0 & 2 & 4 & 1 & 3 & 4 & 0 & 1 & 3 & \cdots \\
\end{array}
\]

Let \( G \) be the outcome subspace and \( X, Y, \) and \( Z \) be the outcome random variables illustrated in Figure 2.8 (page 59). This yields the following statistics:

- geometry of \( G_n \): \( \hat{\zeta}(G_n) = \{0, 1, 2, 3, 4, 5, 6, 7, 8\} \)
- geometry of \( G \): \( \hat{\zeta}(G) = \{4\} \)
- traditional statistics on real line:
  - \( E(X) = \frac{1}{3} \approx 0.333 \)
  - \( \text{Var}(X; E) = \frac{22}{27} \approx 0.815 \)
- outcome subspace statistics on real line:
  - \( \hat{\zeta}(X) = \{\frac{12}{7} \approx 2.4\} \)
  - \( \text{Var}(X; \hat{\zeta}) = \frac{55}{147} \approx 0.376 \)
- outcome subspace statistics on integer line:
  - \( \hat{\zeta}(Y) = \{2, 3\} \)
  - \( \text{Var}(Y; \hat{\zeta}) = \frac{177}{88} \approx 1.982 \)
- outcome subspace statistics on isomorphic structure: \( \hat{\zeta}(Z) = \{4\} \)
  - \( \text{Var}(Z; \hat{\zeta}) = \frac{66}{68} \approx 0.970 \)

Note that unlike the statistics \( E(X) \) and \( \hat{\zeta}(X) \) on the real line, the statistic \( \hat{\zeta}(Z) \) on the isomorphic structure \( K \) yields the maximally likely result, and a much smaller variance as well.

\(^5\) Hull and Dobell (1962), \( \varepsilon \) Jr. and Gentle (1980) page 137 (Theorem 6.1), \( \varepsilon \) Severance (2001) page 86 (Hull-Dobell Theorem)

\(^6\) The sequence \( \{1, 3, 0, 2, 4, 1, 3, 0, 2, 4, 1, 3, \cdots\} \) is generated by the equation \( y_{n+1} = (y_n + 2) \mod 5 \) with \( y_0 = 1 \)
PROOF:

\[ \hat{\mathbf{c}}(\mathbf{G}_k) \triangleq \arg \min \max_{x \in \mathbf{G}_k} d(x, y) \mathbf{P}(y) \]

by definition of \( \hat{\mathbf{c}} \) (Definition 2.3 page 28)

\[ = \arg \min \max_{x \in \mathbf{G}_k} d(x, y) \frac{1}{9} \]

by definition of \( \mathbf{G}_k \)

\[ = \arg \min \max_{x \in \mathbf{G}_k} d(x, y) \]

because \( \phi(x) = \frac{1}{9} x \) is strictly isotone and by Lemma 1.7 page 19

\[ = \arg \min \{4, 4, 4, 4, 4, 4, 4, 4\} \]

because the maximum distance in \( \mathbf{G}_k \) from any \( x \) is 4

\[ = \{0, 1, 2, \ldots, 8\} \]

because the distances for values of \( x \) in \( \mathbf{G}_k \) are the same

\( \hat{\mathbf{c}}(\mathbf{G}) \triangleq \{4\} \)

by weighted ring outcome subspace example (Example 2.6 page 39)

\( \mathbf{E}(X) \triangleq \sum_{x \in \mathbf{R}} x \mathbf{P}(x) \)

by definition of \( \mathbf{E} \) (Definition 1.23 page 9)

\[ = 0 \times \frac{2}{9} + 1 \times \frac{1}{9} + 2 \times \frac{1}{9} + 3 \times \frac{2}{9} + 4 \times \frac{3}{9} = \frac{21}{9} = \frac{7}{3} \approx 2.333 \]

\( \mathbf{Var}(X; \mathbf{E}) = \mathbf{Var}(X) \)

by Theorem 2.1 page 47

\[ \triangleq \sum_{x \in \mathbf{R}} [x - E(X)]^2 \mathbf{P}(x) \]

by definition of \( \mathbf{Var} \) (Definition 1.23 page 9)

\[ = \left(0 - \frac{7}{3}\right)^2 \frac{2}{9} + \left(1 - \frac{7}{3}\right)^2 \frac{1}{9} + \left(2 - \frac{7}{3}\right)^2 \frac{1}{9} + \left(3 - \frac{7}{3}\right)^2 \frac{2}{9} + \left(4 - \frac{7}{3}\right)^2 \frac{3}{9} = \frac{198}{81} = \frac{22}{9} \approx 2.457 \]

\( \hat{\mathbf{E}}(X) \triangleq \arg \min \max_{y \in \mathbf{R}} d(x, y) \mathbf{P}(y) \)

by definition of \( \hat{\mathbf{E}} \) (Definition 2.14 page 46)

\[ = \arg \min \max_{x \in \mathbf{R}} |x - y| \mathbf{P}(y) \]

by definition usual metric on real line (Definition 1.39 page 22)

\[ = \arg \min \begin{cases} |x - 4| \mathbf{P}(4) & \text{for } x \leq \frac{12}{5} \\ |x - 0| \mathbf{P}(0) & \text{otherwise} \end{cases} \]

because expression is minimized at argument \( x = \frac{12}{5} \)

\[ = \left\{ \frac{12}{5} \right\} = 2.4 \]

\( \hat{\mathbf{Var}}(X; \hat{\mathbf{E}}) \triangleq \sum_{x \in \mathbf{R}} d^2(\hat{\mathbf{E}}(X), x) \mathbf{P}(x) \)

by definition of \( \hat{\mathbf{Var}} \) (Definition 2.14 page 46)

\[ = \sum_{x \in \mathbf{R}} d^2 \left(\frac{12}{5}, x\right) \mathbf{P}(x) \]

by \( \hat{\mathbf{E}}(X) \) result

\[ = \left(\frac{12}{5} - 0\right)^2 \frac{2}{9} + \left(\frac{12}{5} - 1\right)^2 \frac{1}{9} + \left(\frac{12}{5} - 2\right)^2 \frac{1}{9} + \left(\frac{12}{5} - 3\right)^2 \frac{2}{9} + \left(\frac{12}{5} - 4\right)^2 \frac{3}{9} \]

\[ = \frac{1}{25 \times 9} \left(2(12 - 0)^2 + 1(12 - 5)^2 + 1(12 - 10)^2 + 2(12 - 15)^2 + 3(12 - 20)^2\right) = \frac{551}{225} \approx 2.449 \]

\( \hat{\mathbf{E}}(Y) \)

\[ \triangleq \arg \min \max_{y \in \mathbf{H}} d(x, y) \mathbf{P}(y) \]

by definition of \( \hat{\mathbf{E}} \) (Definition 2.14 page 46)

\[ = \arg \min \max_{y \in \mathbf{H}} |x - y| \mathbf{P}(y) \]

by definition of integer line (Definition 1.40 page 23)

\[ = \arg \min \max_{y \in \mathbf{H}} \begin{cases} 0 \times 2 & 1 \times 1 & 2 \times 1 & 3 \times 2 & 4 \times 3 \\ 1 \times 2 & 0 \times 1 & 1 \times 1 & 2 \times 2 & 3 \times 3 \\ 2 \times 2 & 1 \times 1 & 0 \times 1 & 1 \times 1 & 2 \times 3 \\ 3 \times 2 & 2 \times 1 & 1 \times 1 & 0 \times 2 & 1 \times 3 \\ 4 \times 2 & 3 \times 1 & 2 \times 1 & 1 \times 2 & 0 \times 3 \end{cases} \]

\[ = \arg \min \begin{cases} \frac{12}{9} & \frac{6}{6} & \frac{8}{8} \end{cases} \]

\[ = \arg \min \begin{cases} \frac{2}{3} \\ \frac{2}{3} \end{cases} \]

\( \hat{\mathbf{Var}}(Y; \hat{\mathbf{E}}) \triangleq \sum_{x \in \mathbf{Z}} d^2(\hat{\mathbf{E}}(Y), x) \mathbf{P}(x) \)

by definition of \( \hat{\mathbf{Var}} \) (Definition 2.14 page 46)

\[ = \sum_{x \in \mathbf{Z}} d^2(2, 3, x) \mathbf{P}(x) \]

by \( \hat{\mathbf{E}}(Y) \) result
\[
E(Z) = Z[\hat{\mathcal{U}}(G)] = \{4\} \quad \text{because } G \text{ and } H \text{ are isomorphic under } Z
\]

by \(\hat{\mathcal{U}}(G)\) result

\[
\text{Var}(Z; \hat{E}) \triangleq \sum_{x \in H} d^2 (E(Z), x) P(x) \quad \text{by definition of } \text{Var} \text{ (Definition 2.14 page 46)}
\]

\[
= \sum_{x \in H} d^2 ((4), x) P(x) \quad \text{by } \hat{E}(Z) \text{ result}
\]

\[
= 1^2 \times \frac{2}{9} + 2^2 \times \frac{1}{9} + 2^2 \times \frac{1}{9} + 1^2 \times \frac{2}{9} + 0^2 \times \frac{3}{9} = \frac{12}{9} = \frac{4}{3} \approx 1.333
\]

Example 2.18 (LCG mappings, sequential ordering).

In Example 2.17 (page 59), the structures \(G_9, G,\) and \(H\) were ordered as a standard ring of integers \((0 < 1 < 2 < \cdots < 7 < 8 < 0 \text{ for } G_9)\). In this current example, as illustrated in Figure 2.9 (page 61), these structures are ordered as they appear in the sequences generated by \(x_{n+1} = (7x_n + 5) \mod 9\) and \(s\) (see Example 2.17 for sequence description). This yields the following statistics:

- geometry of \(G_9\):
  - \(\hat{\mathcal{U}}(G) = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}\)

- geometry of \(G\):
  - \(\hat{\mathcal{U}}(G) = \{3\}\)

- outcome subspace statistics on isomorphic structure: \(E(Z) = \{3\} \quad \text{Var}(Z; \hat{E}) = \frac{10}{9} \approx 1.111\)

Note that a change in ordering structure (from standard ring ordering to sequential ordering) yields a change in statistics \(\hat{\mathcal{U}}(Z) = \{3\} \text{ as opposed to } \hat{E}(Z) = \{4\})\). Intuitively, the sequential ordering of Example 2.18 should yield a better estimate than that of Example 2.17, because it more closely matches the way the PRNG produces a sequence. This intuition is also supported by the variance values \(\text{Var}(Z) = 1\frac{26}{9}\) for standard ring ordering, \(\frac{10}{9}\) for sequential ordering. However, counterintuitively, the sequential ordering no longer yields the maximally likely result of \(\{4\}\).

Proof:

- \(\hat{\mathcal{U}}(G_9) = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}\) by LCG mappings standard ordering example (Example 2.17 page 59)

- \(\hat{\mathcal{U}}(G) = \{3\}\) by Example 2.7 (page 41)

- \(E(Z) = Z[\hat{\mathcal{U}}(G)] = Z[\{3\}]\) by \(\hat{\mathcal{U}}(G)\) result

- \(\text{Var}(Z; \hat{E}) = \text{Var}(G)\) because \(G\) and \(H\) are isomorphic under \(Z\)

- \(= \frac{10}{9} \approx 1.111\) by Example 2.7 (page 41)
Figure 2.10: LCG mappings to linear (X), non-linear discrete (Y) and non-linear continuous (Z) ordered metric spaces (Example 2.19 page 62)

**Example 2.19 (LCG mappings, sequential directed graph).**

Let $G$, $H$ and $Z$ be a illustrated in Figure 2.10 (page 62). In Example 2.18 (page 61), the outcome values were ordered sequentially like a PRNG, but the metrics were commutative, which is unlike a PRNG. In this example, the outcomes are assigned quasi-metrics (Definition D.6 page 153, Remark 1.5 page 22) that are non-commutative. For example in the shaped sequence $s(x, n) = (\ldots, 3, 4, 1, 3, \ldots)$, the “distance” from 3 to 4 is $d(3, 4) = 1$, but from 4 to 3 is $d(4, 3) = 2$. This yields the following statistics:

- geometry of $G_0$: $\hat{z}(G) = \{0, 1, 2, 3, 4, 5, 6, 7, 8\}$
- geometry of $G$: $\hat{z}(G) = \{3\}$

- outcome subspace statistics on isomorphic structure: $\hat{E}(Z) = \{3\}$, $\text{Var}(Z; \hat{E}) = \frac{4}{3} \approx 1.333$

Note that this technique yields the same estimate $\hat{E}(Z) = \{3\}$ as Example 2.18, but with a larger variance.

**Proof:**

\[
\hat{z}(G_0) \triangleq \arg \min_{x \in G_0} \max_{y \in G_0} d(x, y) P(y)
\]

by definition of $\hat{z}$ (Definition 2.3 page 28)

\[
= \arg \min_{x \in G_0} \max_{y \in G_0} d(x, y) \frac{1}{9}
\]

by definition of $G_0$

\[
= \arg \min_{x \in G_0} \max_{y \in G_0} d(x, y)
\]

because $\phi(x) = \frac{1}{9}x$ is strictly isotone and by Lemma 1.7 page 19

\[
= \arg \min \{8, 8, 8, 8, 8, 8, 8, 8, 8\}
\]

because the maximum distance in $G_0$ from any $x$ is 8

\[
= \{0, 1, 2, \ldots, 8\}
\]

because the distances for values of $x$ in $G_0$ are the same

\[
\hat{z}(G) = \{3\}
\]

by Example 2.8 (page 43)

\[
\hat{E}(Z) = Z[\hat{z}(G)]
\]

because $G$ and $H$ are isomorphic under $Z$

\[
= Z[\{3\}]
\]

by $\hat{z}(G)$ result

\[
= \{3\}
\]

by definition of $Z$

\[
\text{Var}(Z; \hat{E}) = \text{Var}(\hat{z}(G))
\]

because $G$ and $H$ are isomorphic under $Z$

\[
= \frac{4}{3} \approx 1.333
\]

by Example 2.8 (page 43)

**Genomic signal processing (GSP) examples**

**Example 2.20 (DNA to linear structures).** Genomic Signal Processing (GSP) analyzes biological sequences called genomes. These sequences are constructed over a set of 4 symbols that are com-
monly referred to as $\mathbb{A}$, $\mathbb{G}$, $\mathbb{C}$, and $\mathbb{G}$, each of which corresponds to a nucleobase (adenine, thymine, cytosine, and guanine, respectively).\footnote{\textcite{Mendel1853} (Mendel (1853): gene coding uses discrete symbols), Watson and Crick (1953a) page 737 (Watson and Crick (1953): gene coding symbols are adenine, thymine, cytosine, and guanine), Watson and Crick (1953b) page 965, Pommerville (2013) page 52} A typical genome sequence contains a large number of symbols (about 3 billion for humans, 29751 for the SARS virus).\footnote{\textcite{GenBank2014} (GenBank (2014) (http://www.ncbi.nlm.nih.gov/genome/guide/human/)(Homo sapiens, NC_000001–NC_000022 (22 chromosome pairs), NC_000023 (X chromosome), NC_000024 (Y chromosome), NC_012920 (mitochondrial)), GenBank (2014) (http://www.ncbi.nlm.nih.gov/nuccore/30271926)(SARS coronavirus, NC_004718.3) S. G. Gregory (2006) (homo sapien chromosome 1), Runtao He (2004) (SARS coronavirus)} Let $G = (\mathbb{A}, \mathbb{G}, \mathbb{C}, d, \varnothing, P)$ be the outcome subspace (Definition 2.1 page 27) generated by a genome where $d$ is the discrete metric (Definition D.8 page 155), $\leq \varnothing$ indicates a completely unordered set (Definition 1.24 page 10), and $P(\mathbb{A}) = P(\mathbb{G}) = P(\mathbb{C}) = P(\varnothing) = 1/4$ (uniformly distributed). Let $H = (\mathbb{R}, |\cdot|, \leq)$ be the real line (Definition 1.39 page 22). This yields the following statistics:

**geometry of $G$:**

$$\hat{\nu}(G) = \{\mathbb{A}, \mathbb{G}, \mathbb{C}, \varnothing\}$$  
**by Example 2.9 page 45**

$$E(X) = \int_{\mathbb{R}} x \mu(x) \, dx$$  
**by definition of $E$ (Definition 1.23 page 9)**

$$= \sum_{x \in \mathbb{C}} x \mu(x)$$  
**by definition of $P$ and Proposition 1.2 page 9**

$$= \frac{1}{4} \sum_{x \in \{0,1,2,3\}} x$$  
**by definitions of $G$, $H$ and $X$**

$$= \frac{1}{4} (0 + 1 + 2 + 3) = \frac{6}{4} = \frac{3}{2} = 1.5$$  
**by Theorem 2.1 page 47**

$$\mathbf{Var}(X; E) = \mathbf{Var}(X)$$  
**by Theorem 2.1 page 47**

$$= \int_{\mathbb{R}} (x - E(X))^2 \mu(x) \, dx$$  
**by definition of $\mathbf{Var}$ (Definition 1.23 page 9)**

$$= \sum_{x \in \mathbb{C}} (x - E(X))^2 \mu(x)$$  
**by definition of $P$ and Proposition 1.2 page 9**

$$= \frac{1}{4} \sum_{x \in \mathbb{H}} \left[ x - \frac{3}{2} \right]^2$$  
**by $E(X)$ result**

Figure 2.11: DNA random variable mappings to real line and integer line (Example 2.20 page 62)
\[
\frac{1}{4} \left[ \left( 0 - \frac{3}{2} \right)^2 + \left( 1 - \frac{3}{2} \right)^2 + \left( 2 - \frac{3}{2} \right)^2 + \left( 3 - \frac{3}{2} \right)^2 \right]
= \frac{1}{4} \cdot 2 \cdot \left[ 0 - 3 \right]^2 + \left( 2 - 3 \right)^2 + (4 - 3)^2 + (6 - 3)^2
= \frac{20}{16} = \frac{5}{4} = 1.25
\]

\[\bar{\mathbb{E}}(X) = \mathbb{E}(X)\] because on real line, \(P\) is symmetric, and by Theorem 2.2 page 47

\[\bar{\mathbb{E}}(X) = \frac{3}{2} = 1.5\] by \(\mathbb{E}(X)\) result

\[\bar{\mathbb{E}}(X) \triangleq \arg \min_{x \in \mathbb{R}} \max_{y \in \mathbb{R}} d(x, y) P(y)\] by definition of \(\bar{\mathbb{E}}\) (Definition 2.14 page 46).  
(alternate proof)

\[\bar{\mathbb{E}}(X) = \arg \min_{x \in \mathbb{R}} \max_{y \in \mathbb{R}} |x - y| P(y)\] by definition usual metric on real line

\[\bar{\mathbb{E}}(X) = \arg \min_{x \in \mathbb{R}} \max_{y \in \{0, 1, 2, 3\}} |x - y| \frac{1}{4}\] by definition of \(G\)

\[\bar{\mathbb{E}}(X) = \arg \min_{x \in \mathbb{R}} \max_{y \in \{0, 1, 2, 3\}} \left\{ \begin{array}{ll} |x - 3| & \text{for } x \leq \frac{3}{2} \\ |x - 0| & \text{otherwise} \end{array} \right\}\] because expression is minimized at argument \(x = \frac{3}{2}\)

\[\bar{\text{Var}}(X; \bar{\mathbb{E}}) = \bar{\text{Var}}(X; \mathbb{E})\] because \(\bar{\mathbb{E}}(X) = \mathbb{E}(X)\)

\[\bar{\text{Var}}(X; \bar{\mathbb{E}}) = \frac{5}{4}\] by \(\bar{\text{Var}}(X; \mathbb{E})\) result

\[\bar{\mathbb{E}}(Y) \triangleq \arg \min_{x \in \mathbb{Z}} \max_{y \in \mathbb{Z}} d(x, y) P(y)\] by definition of \(\bar{\mathbb{E}}\) (Definition 2.14 page 46)

\[\bar{\mathbb{E}}(Y) = \arg \min_{x \in \mathbb{Z}} \max_{y \in \{0, 1, 2, 3\}} |x - y| \frac{1}{4}\] by definition of integer line (Definition 1.40 page 23)

\[\bar{\mathbb{E}}(Y) = \arg \min_{x \in \mathbb{Z}} \max_{y \in \{0, 1, 2, 3\}} |x - y|\]

\[\bar{\text{Var}}(Y; \bar{\mathbb{E}}) \triangleq \sum_{x \in \mathbb{Z}} d^2(\bar{\mathbb{E}}(Y), x) P(x)\] by definition of \(\bar{\text{Var}}\) (Definition 2.14 page 46)

\[\bar{\text{Var}}(Y; \bar{\mathbb{E}}) = \sum_{x \in \mathbb{Z}} d^2([1, 2], x) P(x)\] by \(\bar{\text{E}}(Y)\) result

\[= \left| 0 - 1 \right|^2 \times \frac{1}{4} + \left| 1 - 1 \right|^2 \times \frac{1}{4} + \left| 2 - 2 \right|^2 \times \frac{1}{4} + \left| 3 - 2 \right|^2 \times \frac{1}{4} \times \frac{1}{2} = 0.5\]

\[\Rightarrow\]

Example 2.21 (GSP to complex plane).
A possible solution for the GSP problem (Example 2.20 page 62) is to map \(\mathbb{Z}, \mathbb{R}, \mathbb{Q}, \mathbb{R}\) to the complex plane (Example 1.10 page 23) rather than the real line (Definition 1.39 page 22) such that (see also illustration to the right)

\[a \triangleq X(\mathbb{Z}) = -1 + i\] 
\[t \triangleq X(\mathbb{R}) = 1 + i\]
\[c \triangleq X(\mathbb{Q}) = -1 - i\] 
\[g \triangleq X(\mathbb{R}) = 1 - i\]

However, this solution also is arguably unsatisfactory for two reasons:

1. The order structures are dissimilar. Note that \(c < a\), but \(\mathbb{C}\) and \(\mathbb{Z}\) are incomparable (Definition 1.24 page 10).
2. The metric geometries are dissimilar. Let \( d \) be the \textit{discrete metric} and \( p \) the \textit{usual metric} in \( \mathbb{C} \). Note that
\[
d(X, Y) = d(X, Z) = d(Z, Y) = 1, \quad \text{but} \quad p(a, t) = |a - t| = 2 \neq 2\sqrt{2} = |a - g| = p(a, g).
\]

Example 2.22 (DNA mapping with extended range).
Example 2.20 (page 62) presented a mapping from a DNA structure to a linearly ordered lattices, but the order and metric geometry was not preserved. In this example, a different structure is used that does preserve both order and metric geometry (see illustration to the right). This yields the following statistics:
\[
\hat{E}(X) = \{0\} \quad \hat{\nu}_\sigma(X) = \frac{1}{4}
\]

\begin{proof}
\[
\hat{E}(H) \triangleq \arg \min_{x \in H} \max_{y \in H} d(x, y) \quad P(y) \quad \text{by definition of } \hat{E} \quad \text{(Definition 2.14 page 46)}
\]
\[
= \arg \min_{x \in H[0]} \max_{y \in H[0]} d(x, y) \quad \text{because } P(0) = 0
\]
\[
= \arg \min_{x \in H[0]} \max_{y \in H[0]} \frac{1}{4} \quad \text{by definition of } G
\]
\[
= \arg \min_{x \in H[0]} \max_{y \in H[0]} d(x, y) \quad \text{because } \phi(x) = \frac{1}{4} x \text{ is strictly isotone and by Lemma 1.7 page 19}
\]
\[
= \{0\} \quad \text{because expression is minimized at } x = \{0\}
\]
\[
\hat{\nu}_\sigma(X) \triangleq \sum_{x \in H} d^2(\hat{E}(X), x) \quad P(x) \quad \text{by definition of } \hat{\nu}_\sigma \quad \text{(Definition 2.14 page 46)}
\]
\[
= \sum_{x \in H} d^2(\{0\}, x) \quad P(x) \quad \text{by } \hat{E}(X) \text{ result}
\]
\[
= \sum_{x \in H[0]} \left(\frac{1}{2}\right)^2 \frac{1}{4} = |H \setminus \{0\}| \left(\frac{1}{2}\right)^2 \frac{1}{4} = 4 \left(\frac{1}{2}\right)^2 \frac{1}{4} = \frac{1}{4}
\]
\end{proof}

Example 2.23 (GSP with Markov model). Markov probability models have often been used in genomic signal processing (GSP). A change in the statistics in the sequence may in some cases mean a change in function of the genomic sequence (DNA code). Finding such a change in statistics then is very useful in identifying functions of segments of genomic sequences. Let \( G \) be an outcome subspace (Definition 2.1 page 27) representing a Markov model of depth 2 for a genomic sequence as illustrated in Figure 2.12 (page 66), with joint and conditional probabilities computed over a finite window. Let \( H \) be an outcome subspace isomorphic to \( G \), and \( X \) be a random variable mapping \( G \) to \( H \). A change in the value of the statistic \( \hat{E}(X) \) over the window then may indicate a change in function within the genomic sequence.
2.3 Operations on outcome subspaces

2.3.1 Summation

Example 2.24 (pair of dice outcome subspace). A pair of real dice has a structure as illustrated in Figure 2.13 (page 67). The values represent the standard sum of die faces and thus range from 2 to 12. The table in the figure provides the metric distances between summed values based on the number of edges that must be transversed to move from the first value to the second value. Alternatively, the distance is the number of times the dice must be rotated 90 degrees to move from the first value being face up to the second value being face up. This structure is also illustrated in the undirected graph on the right in Figure 2.13, along with each value's standard probability.

PROOF:

\[ \hat{d}(G) \triangleq \arg \min_{x \in G} \max_{y \in G} d(x, y) P(y) \]

by definition of \( \hat{d} \) (Definition 2.3 page 28)

\[ = \arg \min_{x \in G} \max_{y \in G} \begin{bmatrix} d(2, 2)P(2) & d(2, 3)P(3) & \cdots & d(2, 12)P(12) \\ d(3, 2)P(2) & d(3, 3)P(3) & \cdots & d(3, 12)P(12) \\ \vdots & \vdots & \ddots & \vdots \\ d(12, 2)P(2) & d(12, 3)P(3) & \cdots & d(12, 12)P(12) \end{bmatrix} \]

\( ^* \)Many many thanks to Katie L. Greenhoe and Jonathan J. Greenhoe for help computing these values.
Figure 2.13: metrics based on number of die edges for a pair of real dice (Example 2.24 page 66)
\[
\sum_{x \in \mathbb{G}} [d(7, x)]^2 \mathbb{P}(x) = \frac{1}{36} (2^2 \times 1 + 1^2 \times 2 + 1^2 \times 3 + 1^2 \times 4 + 1^2 \times 5 + 0^2 \times 6 + 1^2 \times 5 + 1^2 \times 4 + 1^2 \times 3 + 1^2 \times 2 + 2^2 \times 1)
\]
\[
= \frac{1}{36} (4 + 2 + 3 + 4 + 5 + 0 + 5 + 4 + 3 + 2 + 4)
\]
\[
= \frac{36}{36} = 1
\]

by definition of \(\mathbb{V}_x\) (Definition 2.14 page 46)

The next two examples are examples of sums of \textit{outcome subspaces} (Definition 2.1 page 27): Example 2.25 page 68 (sum of dice pair) and Example 2.27 page 71 (sum of spinner pair).

\[\text{Figure 2.14: pair of dice mappings (Example 2.14 page 54)}\]

\textbf{Example 2.25 (pair of dice and hypothesis testing).} Let \(\mathbf{G}\) be the \textit{pair of dice outcome subspace} (Example 2.24 page 66), \(X \in (\mathbb{R}, |\cdot|, \leq)^\mathbb{G}\) an \textit{outcome random variable} mapping from \(\mathbf{G}\) to the \textit{real line} (Definition 1.39 page 22), and \(X \in H^\mathbb{G}\) a mapping to a structure \(H\) that is \textit{isomorphic} to \(\mathbf{G}\), as illustrated in Figure 2.14 (page 68). This yields the following statistics:

\[
\mathbb{G} \quad \hat{\mathbb{G}}(\mathbf{G}) = \hat{\mathbb{G}}_G(\mathbf{G}) = \{7\} \quad \mathbb{V}_x(\mathbf{G}) = 1
\]

\[
\text{traditional statistics on real line } (\mathbb{R}, |\cdot|, \leq): \quad \hat{E}(X) = 7 \quad \mathbb{V}_x(W; \hat{E}) = \frac{35}{6} \approx 5.833
\]

\[
\text{outcome subspace statistics on real line } (\mathbb{R}, |\cdot|, \leq): \quad \hat{E}(X) = \{7\} \quad \mathbb{V}_x(W; \hat{E}) = \frac{35}{6} \approx 5.833
\]

\[
\text{outcome subspace statistics on isomorphic structure } H: \quad \hat{E}(Y) = \{7\} \quad \mathbb{V}_x(Y; \hat{E}) = 1
\]

Although the expected values of both \textit{outcome subspaces} are the essentially the same (7 and \{7\}), the isomorphic structure \(H\) yields a much smaller variance (a much smaller expected error). This is significant in statistical applications such as hypothesis testing. Suppose for example we have two \textit{pair of real dice} (Example 2.2 page 31), one pair being made of two uniformly distributed die and one pair of weighted die. We want to know which pair is the uniform die. So we roll each pair one time. Suppose the outcome of the first pair is 11 and the the outcome of the second pair is 6. Which pair is more likely to be the uniform pair? Using traditional statistical analysis, the answer is the second pair, because it is closer to the expected value (0.414 standard deviations as opposed to 1.656 standard deviations). However, this result is deceptive, because as can be seen in Figure 2.13 (page 67), the distance from the expected value to the values 11 and 6 are the same (\(d(7, 11) = d(7, 6) = 1 = 1\) standard deviation). So arguably the outcome of the single roll test would contribute nothing to a good decision algorithm.

\% PROOF:

\[
\hat{\mathbb{G}}(\mathbf{G}) = \{7\} \quad \text{by Example 2.24 (page 66)}
\]

\[
\mathbb{V}_x(\mathbf{G}) = 1 \quad \text{by Example 2.24 (page 66)}
\]
E(X) \triangleq \int_{x \in \mathbb{R}} xP(x) \, dx \quad \text{by definition of } E \text{ (Definition 1.23 page 9)}
\triangleq \frac{1}{36} \sum_{x \in \mathbb{Z}} x36P(x) \quad \text{by definition of } P
= \frac{1}{36} (2 \times 1 + 3 \times 2 + 4 \times 3 + 5 \times 4 + 6 \times 5 + 7 \times 6 + 8 \times 5 + 9 \times 4 + 10 \times 3 + 11 \times 2 + 12 \times 1)
= \frac{252}{36} = 7
\text{Var}(X; E) = \text{Var}(X) \quad \text{by Theorem 2.1 page 47}
= \int_{x \in \mathbb{R}} [x - E(X)]^2 P(x) \, dx \quad \text{by definition of } \text{Var} \text{ (Definition 1.23 page 9)}
= \sum_{x \in \mathbb{Z}} [x - E(X)]^2 P(x) \quad \text{by definition of } P
= \sum_{x \in \mathbb{Z}} (x - 7)^2 \frac{1}{36} \quad \text{by } E(X) \text{ result}
= 2 \frac{25 \times 1 + 16 \times 2 + 9 \times 3 + 4 \times 4 + 1 \times 5}{36} = \frac{35}{6} \approx 5.833
\bar{E}(X) = E(X) \quad \text{by Theorem 2.2 (page 47)}
= \{7\} \quad \text{by } E(X) \text{ result}
\text{Var}(X; \bar{E}) = \text{Var}(X; E) \quad \text{because } \bar{E}(X) \equiv E(X)
= \frac{35}{6} \approx 5.833 \quad \text{by } \text{Var}(X; E) \text{ result}
\bar{E}(Y) = Y[\hat{\mathcal{U}}(G)] \quad \text{because } G \text{ and } H \text{ are isometric under } Y
= Y(\{7\}) \quad \text{by } \hat{\mathcal{U}}(G) \text{ result}
= \{7\} \quad \text{by definition of } Y
\text{Var}(Y; \bar{E}) = \text{Var}(G) \quad \text{because } G \text{ and } H \text{ are isometric under } Y
= 1 \quad \text{by } \text{Var}(G) \text{ result}

\begin{center}
\begin{tikzpicture}
  \node[draw] at (0,0) {\textbf{spinner} (Example 2.4 page 35)};
  \node[draw] at (2,0) {\textbf{spinner} (Example 2.4 page 35)};
  \node[draw] at (4,0) {\textbf{spinner} (Example 2.6 page 70)};
\end{tikzpicture}
\end{center}

Figure 2.15: metrics based on a pair of spinners (Example 2.26 page 70)
Example 2.26 (pair of spinners). A pair of spinners (Example 2.4 page 35) has a structure as illustrated in Figure 2.15 (page 69). The values represent the standard sum of spinner positions \(1, 2, \ldots, 6\) and thus range from 2 to 12. The table in the figure provides the metric distances between summed values based on how many positions one must traverse to get from one value to the next (in which ever direction is shortest). This structure is also illustrated in the undirected graph in the upper right of Figure 2.15, along with each value's standard probability.

**Proof:**

\[
\hat{\mathcal{C}}(G) = \arg \min_{x \in G} \max_{y \in G} d(x, y) P(y) \quad \text{by definition of } \hat{\mathcal{C}} \text{ (Definition 2.3 page 28)}
\]

\[
= \arg \min_{x \in G} \max_{y \in G} \frac{1}{36} \begin{pmatrix}
0 & 2 & 6 & 12 & 10 & 6 & 10 & 12 & 12 & 6 & 2 \\
1 & 0 & 3 & 8 & 15 & 12 & 5 & 8 & 9 & 6 & 4 \\
2 & 2 & 0 & 4 & 10 & 18 & 10 & 4 & 6 & 6 & 4 \\
3 & 3 & 3 & 0 & 5 & 12 & 15 & 8 & 3 & 4 & 3 \\
2 & 6 & 6 & 4 & 0 & 6 & 10 & 12 & 6 & 2 & 2 \\
1 & 4 & 9 & 8 & 5 & 0 & 5 & 8 & 9 & 4 & 1 \\
2 & 2 & 6 & 12 & 10 & 6 & 0 & 4 & 6 & 6 & 2 \\
3 & 4 & 3 & 8 & 15 & 12 & 5 & 0 & 3 & 4 & 3 \\
4 & 6 & 6 & 4 & 10 & 18 & 10 & 4 & 0 & 2 & 2 \\
3 & 8 & 9 & 8 & 5 & 12 & 15 & 8 & 3 & 0 & 1 \\
2 & 6 & 12 & 12 & 10 & 6 & 10 & 12 & 6 & 2 & 0
\end{pmatrix}
\]

\[
\hat{\mathcal{C}}_q(G) = \arg \min_{x \in G} \sum_{y \in G} d(x, y) P(y) \quad \text{by definition of } \hat{\mathcal{C}}_q \text{ (Definition 2.4 page 28)}
\]

\[
= \arg \min_{x \in G} \frac{1}{36} \begin{pmatrix}
0 + 2 & 6 + 12 & 10 & 6 + 10 & 12 + 6 & 12 + 6 & 2 \\
1 + 0 + 3 & 8 + 15 + 12 + 5 & 8 + 9 + 8 + 3 \\
2 + 2 + 0 & 4 + 10 + 18 + 10 & 4 & 6 + 6 & 4 + 6 \\
3 + 4 + 3 & 0 + 5 & 12 + 15 & 8 & 3 & 4 + 3 \\
2 + 6 + 6 & 4 & 0 + 6 & 10 + 12 & 6 & 2 + 2 \\
1 + 4 + 9 & 8 & 5 & 0 & 5 & 8 & 9 + 4 + 1 \\
2 + 2 + 6 & 12 + 10 & 6 & 0 + 4 & 6 + 6 & 4 + 2 \\
3 + 4 + 3 & 8 + 15 + 12 & 5 & 0 & 3 & 4 + 3 \\
4 + 6 + 6 & 4 & 10 + 18 & 10 & 4 & 0 + 2 + 2 \\
3 + 8 + 9 & 8 & 5 & 12 & 15 & 8 & 3 & 0 + 1 \\
2 + 6 + 12 & 12 + 10 & 6 & 10 + 12 & 6 & 2 + 0
\end{pmatrix}
\]

\[
= \arg \min_{x \in G} \frac{1}{36} \begin{pmatrix}
78 \\
72 \\
66 \\
60 \\
56 \\
54 \\
56 \\
60 \\
66 \\
72 \\
78
\end{pmatrix}
\]

\[
\hat{\mathcal{V}}_{ar}(G) = \sum_{x \in G} [d(\hat{\mathcal{C}}(G), x)]^2 P(x) \quad \text{by definition of } \hat{\mathcal{V}}_{ar} \text{ (Definition 2.14 page 46)}
\]

\[
= \sum_{x \in G} d(7, x)^2 P(x) \quad \text{by definition of } \hat{\mathcal{V}}_{ar} \text{ (Definition 2.14 page 46)}
\]

\[
= \frac{1}{36} \left( 1^2 1 + 2^2 1 + 3^2 1 + 2^2 4 + 1^2 5 + 0^2 6 + 1^2 5 + 2^2 4 + 2^2 3 + 2^2 2 + 1^2 1 \right)
\]

\[
= \frac{1}{36} \left( 2 + 8 + 27 + 16 + 5 + 0 + 5 + 16 + 12 + 8 + 1 \right)
\]

\[
= \frac{100}{36} = \frac{25}{9} = 2.778 \approx 2.778
\]

\[\]
**Example 2.27** (pair of spinner and hypothesis testing). Let $G$ be a pair of spinners (Example 2.26 page 70), $X$ a random variable mapping to the real line (Definition 1.39 page 22), and $Y$ a random variable mapping to an ordered metric space (Definition 1.38 page 22) that is isomorphic to $G$ under $Y$, as illustrated in Figure 2.16 (page 71). This yields the following statistics:

- **Geometry of $G$:**
  - $\hat{\mathcal{C}}(G) = \hat{\mathcal{C}}_0(G) = \{7\}$
  - $\text{Var}(G) = \frac{25}{9} \approx 2.778$

- **Traditional statistics on real line ($\mathbb{R}, |\cdot|, \leq$):**
  - $E(X) = 7$
  - $\text{Var}(W) = \frac{35}{3} \approx 5.833$

- **Outcome subspace statistics on real line ($\mathbb{R}, |\cdot|, \leq$):**
  - $E(X) = 7$
  - $\text{Var}(W; \hat{E}) = \frac{35}{3} \approx 5.833$

- **Outcome subspace statistics on isomorphic structure $H$:**
  - $E(Y) = \hat{\mathcal{C}}(G) = \{7\}$
  - $\text{Var}(Y; \hat{E}) = 1$

Although the expected value of both outcome subspaces are the same ($E(X) = E(Y) = 7$), the isomorphic outcome subspace $H$ yields a much smaller variance (a much smaller expected error). This is significant in statistical applications such as hypothesis testing. Suppose for example we have two pair of spinners (Example 2.4 page 35), one pair being made of two uniformly distributed spinners, and one pair of weighted spinners. We want to estimate which is which. So we spin each pair one time. Suppose the outcome of the first pair is 12 and the the outcome of the second pair is 10. Which pair is more likely to be the uniform pair? Using traditional statistical analysis, the answer is the second pair, because it is closer to the expected value (true $d(7, 10) = |7 - 10| = 3$ is 1.8 standard deviations as opposed to $d(7, 12) = |7 - 12| = 5$ = 3 standard deviations). However, this result is deceptive, because as can be seen in the table in Figure 2.15 (page 69), 12 is actually closer to the expected value in $G$ than is 10 ($d(7, 12) = 1 < 3 = d(7, 10)$). So arguably the better choice, based on this one trial, is the first pair.

**Proof:**

- $\hat{\mathcal{C}}(G) = \{7\}$ by Example 2.26 (page 70)
- $\hat{\mathcal{C}}_0(G) = \{7\}$ by Example 2.26 (page 70)
- $\text{Var}(G) = \frac{25}{9} \approx 2.778$ by Example 2.26 (page 70)
- $E(X) = \frac{252}{36} = 7$ by Example 2.25 (page 68)
- $\text{Var}(X) = \frac{35}{6} \approx 5.833$ by Example 2.25 (page 68)
- $E(X) = E(X)$ because on real line, $P$ is symmetric, and by Theorem 2.2 page 47
- $\text{Var}(X; E) = \text{Var}(X; \hat{E})$ because $E(X) = E(X)$
- $\text{Var}(X)$ by Theorem 2.1 page 47
- $\frac{35}{6} \approx 5.833$
- $\hat{E}(Y) = Y[\hat{\mathcal{C}}(G)]$ because $G$ and $H$ are isomorphic under $Y$
2.3. OPERATIONS ON OUTCOME SUBSPACES

\[
\begin{align*}
\var(Y; \bar{E}) &= \var(Y) \\
&= \var(\mathcal{G}) \\
&= 1 \\
\end{align*}
\]

by Example 2.26 (page 70)

because \( \mathcal{G} \) and \( H \) are isomorphic under \( Y \)

by Example 2.26 (page 70)

Figure 2.17: real line addition arg \( \min_x \max_y \) calculation graph (Example 2.28 page 72)

**Example 2.28** (linear addition). Let \( X \) be a random variable (Definition 2.13 page 46) mapping to a real line ordered metric space (Definition 1.39 page 22) resulting in probability values of

\[ P(0) = P(2) = \frac{1}{4}, \quad P(x) = 0 \text{ otherwise.} \]

Let \( Y \) be a random variable mapping to a real line ordered metric space resulting in probability values of

\[ P(0) = P(1) = \frac{1}{4}, \quad P(2) = \frac{1}{2}, \quad P(x) = 0 \text{ otherwise.} \]

Let \( Z \triangleq X + Y \) be the random variable mapping to the outcome subspace (Definition 2.1 page 27) induced by adding \( X \) and \( Y \) resulting in probabilities

\[
\begin{array}{c|cccc}
0 & 1 & 2 & 3 & 4 \\
\hline
\frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} & \frac{1}{8} \\
\end{array}
\]

and \( P(z) = 0 \text{ otherwise.} \)

Note that although the traditional expectation \( \mathbb{E} \) (Definition 1.23 page 9) distributes over addition such that

\[
\mathbb{E}(X + Y) = \frac{18}{8} = 1 + \frac{5}{4} = \mathbb{E}(X) + \mathbb{E}(Y),
\]

the alternative expectation \( \hat{E} \) (Definition 2.14 page 46) does **not**:

\[
\hat{E}(X + Y) = \frac{5}{8} \neq \frac{7}{4} = \hat{E}(X) + \hat{E}(Y).
\]

**Proof:**

\[
\begin{align*}
\mathbb{E}(X) &= \int_{x \in \mathbb{R}} xP(x) \, dx \\
&= \sum_{x \in \mathbb{Z}} xP(x) \, dx \\
&= 0 \times \frac{1}{2} + 2 \times \frac{1}{2} \\
&= 1 \\
\end{align*}
\]

by definition of \( \mathbb{E} \) (Definition 1.23 page 9)

by definition of \( P \) and Proposition 1.2 page 9

by definition of \( P \)

\[
\begin{align*}
\mathbb{E}(Y) &= \int_{y \in \mathbb{R}} yP(y) \, dy \\
&= \sum_{y \in \mathbb{Z}} yP(y) \, dy \\
&= 0 \times \frac{1}{4} + 1 \times \frac{1}{4} + 2 \times \frac{1}{2} \\
&= \frac{5}{4} \\
\end{align*}
\]

by definition of \( \mathbb{E} \) (Definition 1.23 page 9)

by definition of \( P \) and Proposition 1.2 page 9

by definition of \( P \)

\[
\begin{align*}
\mathbb{E}(Z) &= \int_{z \in \mathbb{R}} zP(z) \, dz \\
&= \sum_{z \in \mathbb{Z}} zP(z) \\
&= \frac{5}{4} \\
\end{align*}
\]

by definition of \( \mathbb{E} \) (Definition 1.23 page 9)

by definition of \( P \) and Proposition 1.2 page 9
\[ = 0 \times \frac{1}{8} + 1 \times \frac{1}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} + 4 \times \frac{2}{8} \] by definition of \( P \)

\[ = \frac{9}{4} \]

\[ \mathcal{E}(X) = \arg \min_{x \in \mathbb{R}} \max_{y \in \mathbb{R}} d(x, y) P(y) \] by definition of \( \mathcal{E} \) (Definition 2.14 page 46)

\[ = \arg \min_{x \in \mathbb{Z}} \max_{y \in \mathbb{Z}} d(x, y) P(y) \] 

\[ = \arg \min_{x \in \mathbb{Z}} \left\{ \begin{array}{ll}
|2 - x| \frac{1}{2} & \text{for } x \leq 1 \\
|x| \frac{1}{2} & \text{otherwise}
\end{array} \right. \]

\[ = \{1\} \] 

because \( \max(x) \) is minimized at \( x = 1 \)

\[ \mathcal{E}(Y) = \arg \min_{x \in \mathbb{R}} \max_{y \in \mathbb{R}} d(x, y) P(y) \] by definition of \( \mathcal{E} \) (Definition 2.14 page 46)

\[ = \arg \min_{x \in \mathbb{Z}} \max_{y \in \mathbb{Z}} d(x, y) P(y) \] by definition of \( P(x) \)

\[ = \arg \min_{x \in \mathbb{Z}} \left\{ \begin{array}{ll}
|2 - x| \frac{1}{2} & \text{for } \frac{4}{3} \geq x \geq 4 \\
|x| \frac{1}{2} & \text{otherwise}
\end{array} \right. \]

\[ = \left\{ \frac{4}{3} \right\} \] 

because \( \max(y) \) is minimized at \( y = \frac{4}{3} \)

\[ \mathcal{E}(Z) = \arg \min_{x \in \mathbb{R}} \max_{y \in \mathbb{R}} d(x, y) P(y) \] by definition of \( \mathcal{E} \) (Definition 2.14 page 46)

\[ = \arg \min_{x \in \mathbb{Z}} \max_{y \in \mathbb{Z}} d(x, y) P(y) \] by definition of \( P(x) \)

\[ = \arg \min_{x \in \mathbb{Z}} \left\{ \begin{array}{ll}
|x - 2| \frac{1}{3} & \text{for } -2 \geq x \geq 3 \\
|x - 4| \frac{1}{3} & \text{for } -2 \leq x \leq \frac{8}{3} \\
|x| \frac{1}{8} & \text{for } \frac{8}{3} \leq x \leq 3
\end{array} \right. \]

\[ = \left\{ \frac{8}{3} \right\} \] 

because \( \max(z) \) is minimized at \( z = \frac{8}{3} \) (see Figure 2.17 page 72)

\[ \Rightarrow \]

### 2.3.2 Multiplication

**Example 2.29** (ring multiplication). Let \( X \in H^G \) be a random variable where \( G \) is the *weighted spinners* illustrated in Figure 2.18 (page 73). Note that, in agreement with Corollary 2.1 (page 48),

\[ \mathcal{E}(2X) = \{4\} = \{2 \times 5 \mod 6\} = 2\{5\} \mod 6 = 2\mathcal{E}(X) \mod 6 . \]

**Proof:**

\[ \hat{\mathcal{C}}(G) \triangleq \arg \min_{x \in H} \max_{y \in H} d(x, y) P(y) \] by definition of \( \hat{\mathcal{C}} \) (Definition 2.3 page 28)
2.3. OPERATIONS ON OUTCOME SUBSPACES

$$\begin{align*}
&= \arg \min_{x \in H} \max_{y \in H} \frac{1}{10} \begin{bmatrix}
0 \times 1 & 1 \times 1 & 1 \times 1 & 3 \times 1 & 2 \times 1 & 1 \times 5 \\
1 \times 1 & 0 \times 1 & 1 \times 1 & 2 \times 1 & 3 \times 1 & 2 \times 5 \\
2 \times 1 & 1 \times 1 & 0 \times 1 & 2 \times 1 & 3 \times 1 & 3 \times 5 \\
3 \times 1 & 2 \times 1 & 1 \times 1 & 0 \times 1 & 1 \times 1 & 0 \times 5 \\
2 \times 1 & 3 \times 1 & 2 \times 1 & 1 \times 1 & 0 \times 1 & 1 \times 5 \\
1 \times 1 & 2 \times 1 & 3 \times 1 & 2 \times 1 & 1 \times 1 & 0 \times 5 
\end{bmatrix} = \arg \min_{x \in H} \frac{1}{10} \begin{bmatrix}
5 \\
10 \\
15 \\
10 \\
5 \\
3 
\end{bmatrix} \\
&= \{5\} \\
\hat{G}(G) &\triangleq \arg \min_{x \in H} \sum_{y \in H} d(x, y) P(y) \quad \text{by definition of } \hat{G} \quad \text{(Definition 2.4 page 28)} \\
&= \arg \min_{x \in H} \frac{1}{10} \begin{bmatrix}
0 \times 1 + 1 \times 1 + 2 \times 1 + 3 \times 1 + 2 \times 1 + 1 \times 5 \\
1 \times 1 + 0 \times 1 + 1 \times 1 + 2 \times 1 + 3 \times 1 + 2 \times 5 \\
3 \times 1 + 2 \times 1 + 1 \times 1 + 0 \times 1 + 1 \times 1 + 2 \times 5 \\
2 \times 1 + 3 \times 1 + 2 \times 1 + 1 \times 1 + 0 \times 1 + 1 \times 5 \\
1 \times 1 + 2 \times 1 + 3 \times 1 + 2 \times 1 + 1 \times 1 + 0 \times 5 
\end{bmatrix} = \arg \min_{x \in H} \frac{1}{10} \begin{bmatrix}
13 \\
17 \\
21 \\
17 \\
9 
\end{bmatrix} \\
&= \{5\} \\
\hat{E}(X) &= X[\hat{G}(G)] \\
&= X[\{5\}] \quad \text{by } \hat{G}(G) \text{ result} \\
&= \{5\} \\
\hat{E}(2X) &\triangleq \arg \min_{x \in H} \max_{y \in H} d(x, y) P(y) \quad \text{by definition of } \hat{E} \quad \text{(Definition 2.14 page 46)} \\
&= \arg \min_{x \in H} \max_{y \in H} \frac{1}{10} \begin{bmatrix}
0 \times 2 & 1 \times 2 & 1 \times 6 \\
1 \times 2 & 0 \times 2 & 1 \times 6 \\
1 \times 2 & 1 \times 2 & 0 \times 6 
\end{bmatrix} = \arg \min_{x \in H} \frac{1}{10} \begin{bmatrix}
6 \\
6 \\
2 
\end{bmatrix} \\
&= \{4\}
\end{align*}$$

2.3.3 Metric transformation

It is possible to use a metric transform (Definition D.11 page 156) to transform the structure of an outcome subspace (Definition 2.1 page 27) into a completely different outcome subspace. This is demonstrated in Example 2.30 (page 75)—Example 2.32 (page 76). Naturally, by doing so one can sometimes even change the geometric centers (Definition 2.3 page 28, Definition 2.4 page 28) of the outcome subspaces, and hence also the statistics of random variables that map to/from them. This is demonstrated in Example 2.32 (page 76)—Example 2.33 (page 76).

**Theorem 2.4.** Let $\phi$ be a metric preserving function (Definition D.11 page 156). Let $G \triangleq (\Omega, \preceq, d, \hat{P})$ and $H$ be outcome subspaces (Definition 2.1 page 27).

$\left\{\begin{array}{c}
(1). \quad \phi(H) = G \\
(2). \quad \phi \text{ is strictly isotone and } \hat{P} \text{ is uniform } \\
(3). \quad \end{array}\right\} \implies \hat{C}(H) = \hat{C}(G)$

**Proof:**

$$\begin{align*}
\hat{C}(H) &= \phi[\hat{C}(H)] \\
&= \arg \min_{x \in G} \max_{y \in G} \phi[d(x, y)] P(y) \\
&= \arg \min_{x \in G} \max_{y \in G} d(x, y) \\
&= \hat{C}(G) \\
&= \hat{C}(G) \\
&= \hat{C}(G)
\end{align*}$$

by hypothesis (1)

by definition of $\hat{C}$ (Definition 2.3 page 28)

by hypothesis (3) and Lemma 1.7 page 19

by hypothesis (2) and Lemma 1.7 page 19

by definition of $\hat{C}$ (Definition 2.3 page 28)
Example 2.30 (discrete metric transform on outcome subspaces). Let \( g \) be a function (a pullback function Theorem 1.7 page 21) such that \( g(\Box) = \Box, g(\Diamond) = \Diamond, g(\bigstar) = \bigstar, \) and \( g(\bigcirc) = \bigcirc. \) Then under the discrete metric preserving function \( \phi \) (Example D.7 page 158) the real die outcome subspace (Example 2.2 page 31) becomes the fair die outcome subspace (Example 2.1 page 29), and under \( \phi \circ g \) the spinner outcome subspace (Example 2.4 page 35) also becomes the fair die outcome subspace, as illustrated in Figure 2.19 (page 75). This yields the following geometric statistics:

\[
\hat{C}(G) = \hat{C}(H) = \{\Box, \bigstar, \Diamond, \bigcirc, \}.
\]

Example 2.31. Let \( \phi_1 \) be the metric preserving function defined in Example D.11 (page 158). Then under \( \phi_1, \) the spinner outcome subspace (Example 2.4 page 35) becomes what is here called the wagon wheel output subspace, as illustrated on the left in Figure 2.20 (page 75). Let \( G \) be the spinner outcome subspace and \( H \) the wagon wheel outcome subspace. This yields the following geometric statistics:

\[
\hat{C}(G) = \{\Box, \bigcirc\} \quad \hat{C}(H) = \{\bigcirc\}.
\]

Note that the metric transform \( \phi_1 \) also moves the outcome center from one that is not maximally likely, to one that is.

**Proof:**

\[
\hat{C}(G) \triangleq \arg \min_{x \in G} \max_{y \in G} d(x, y) P(y) \quad \text{by definition of } \hat{C} \text{ (Definition 2.3 page 28)}
\]

\[
= \arg \min_{x \in G} \max_{y \in G} \frac{1}{10} \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 1 \\ 1 & 0 & 2 & 1 & 3 & 4 \\ 2 & 1 & 0 & 1 & 2 & 4 \\ 3 & 1 & 2 & 1 & 1 & 2 \\ 2 & 1 & 3 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 8 & 12 & 4 \\ 14 & 6 & 8 \\ 14 & 4 & 6 \\ 14 & 6 & 8 \\ 14 & 6 & 8 \end{bmatrix} = \{\bigcirc\}
\]

\[
\hat{C}(H) \triangleq \arg \min_{x \in G} \max_{y \in G} d(x, y) P(y) \quad \text{by definition of } \hat{C} \text{ (Definition 2.3 page 28)}
\]

\[
= \arg \min_{x \in G} \max_{y \in G} \frac{1}{10} \begin{bmatrix} 0 & 1 & 2 & 3 & 4 & 1 \\ 1 & 0 & 2 & 1 & 3 & 4 \\ 2 & 1 & 0 & 1 & 2 & 4 \\ 3 & 1 & 2 & 1 & 1 & 2 \\ 2 & 1 & 3 & 1 & 1 & 1 \\ 1 & 1 & 2 & 3 & 1 & 1 \end{bmatrix} \begin{bmatrix} 8 & 12 & 4 \\ 14 & 6 & 8 \\ 14 & 4 & 6 \\ 14 & 6 & 8 \\ 14 & 6 & 8 \end{bmatrix} = \{\bigcirc\}
\]
2.3. OPERATIONS ON OUTCOME SUBSPACES

Example 2.32. Let $\phi$ be the metric preserving function defined in Example D.8 (page 158). Let $g$ be the function defined in Example 2.30 (page 75). Then under $\phi \ast g$, the spinner outcome subspace (Example 2.4 page 35) becomes the weighted die outcome subspace (Example 2.3 page 33), as illustrated on the right in Figure 2.20 (page 75). Let $G$ be the spinner outcome subspace and $H$ the weighted die outcome subspace. These structures have the following geometric statistics:

$$\hat{G}(G) = \{4, \bar{6}\} \quad \hat{G}(H) = \{1, 3, 4, \bar{5}, \bar{6}\}.$$

Note that in Example 2.31 (page 75), the metric transform $\phi_1$ results in a smaller (smaller cardinality Definition 1.13 page 6) center $|\hat{G}(G)| = 2 > 1 = |\hat{G}(H)|$. But here, the metric transform $\phi_2 \ast g$ results in a larger center. $|\hat{G}(G)| = 2 < 5 = |\hat{G}(H)|$.

Proof:

$$\hat{G}(G) = \{4, \bar{6}\} \quad \text{by Example 2.31 (page 75)}$$

$$\hat{G}(H) = \arg\min_{x \in H} \max_{y \in G} d(x, y) P(y) \quad \text{by definition of } \hat{G} \text{ (Definition 2.3 page 28)}$$

$$\hat{G}(H) = \arg\min_{x \in H} \max_{y \in G} \frac{1}{10} \left\{ \begin{array}{cccccccc} 0 \times 1 & 1 \times 2 & 2 \times 1 & 1 \times 1 & 1 \times 4 & 2 \times 1 & \bar{1} \times 1 & \bar{1} \times 2 \\
1 \times 1 & 0 \times 2 & 1 \times 1 & 1 \times 1 & 2 \times 4 & 1 \times 1 & \bar{2} \times 1 & \bar{1} \times 2 \\
\bar{1} \times 1 & \bar{1} \times 2 & 0 \times 1 & 2 \times 1 & 1 \times 4 & 1 \times 1 & \bar{1} \times 1 & \bar{2} \times 1 \\
\bar{2} \times 1 & \bar{2} \times 1 & \bar{1} \times 1 & 1 \times 1 & 0 \times 4 & 1 \times 1 & \bar{2} \times 1 & \bar{1} \times 1 \\
\end{array} \right\} = \arg\min_{x \in H} \frac{1}{10} \left\{ \begin{array}{c} 8 \\
4 \\
4 \\
4 \\
\end{array} \right\} = \{1\} \quad \text{by Example 2.31 (page 75)}$$

$$\hat{G}(H) = \arg\min_{x \in H} \max_{y \in G} \frac{1}{10} \left\{ \begin{array}{cccccccc} 0 \times 1 & 1 \times 2 & 2 \times 1 & 1 \times 1 & 1 \times 4 & 2 \times 1 & \bar{1} \times 1 & \bar{1} \\
1 \times 1 & 0 \times 2 & 1 \times 1 & 1 \times 1 & 2 \times 4 & 1 \times 1 & \bar{2} \times 1 & \bar{1} \\
\bar{1} \times 1 & \bar{1} \times 2 & 0 \times 1 & 2 \times 1 & 1 \times 4 & 1 \times 1 & \bar{1} \times 1 & \bar{2} \\
\bar{2} \times 1 & \bar{2} \times 1 & \bar{1} \times 1 & 1 \times 1 & 0 \times 4 & 1 \times 1 & \bar{2} \times 1 & \bar{1} \\
\end{array} \right\} = \arg\min_{x \in H} \frac{1}{10} \left\{ \begin{array}{c} 8 \\
4 \\
4 \\
4 \\
\end{array} \right\} = \{1\} \quad \text{by Example 2.31 (page 75)}$$

Figure 2.21: real line addition arg min$_x$ max$_y$ calculation graph (Example 2.33 page 76)

Example 2.33 (linear addition with metric transform). Example 2.28 (page 72) gave the result $\hat{E}(X + Y) = \frac{5}{7}$ rather than the perhaps more desirable result of $\frac{7}{7}$ (which equals $1 + \frac{2}{7} = \hat{E}(X) + \hat{E}(Y)$). Again, we adjust the geometric statistics of an outcome subspace by use of a metric preserving function (Definition D.11 page 156). In particular, we use the power transform/snowflake transform (Example D.4 page 158) $f(x) = x^a$. If we let $a = \frac{\ln 2}{\ln 1 - \ln 3} \approx 2.0600427$, then $\hat{E}(X + Y) = \frac{7}{7}$, as illustrated in Figure 2.21 (page 76).
CHAPTER 3

SYMBOLIC SEQUENCE PROCESSING ON R^N

"...those who assert that the mathematical sciences say nothing of the beautiful, or the good are in error. For these sciences say and prove a great deal about them; if they do not expressly mention them, but prove attributes which are their results or definitions, it is not true that they tell us nothing about them. The chief forms of beauty are order and symmetry and definiteness, which the mathematical sciences demonstrate in a special degree."

Aristotle 384 BC – 322 BC, Greek philosopher

3.1 Outcome subspace sequences

3.1.1 Definitions

Definition 3.1. Let D_1 and D_2 be convex subsets (Definition 1.29 page 11) of Z. Let D \triangleq (\bigwedge D_1 - \bigvee D_2 - 1 : \bigvee D_1 - \bigwedge D_2 + 1). Let \( \langle x_n \rangle_{D_1} \) and \( \langle y_n \rangle_{D_2} \) be sequences over an outcome subspace \((\Omega, \leq, d, \hat{P})\). The outcome subspace sequence metric \( p(\langle x_n \rangle, \langle y_n \rangle) \) is defined as

\[
p(\langle x_n \rangle, \langle y_n \rangle) \triangleq \sum_{n \in D} f(n) \text{ where } f(n) \triangleq \begin{cases} d(x_n, y_n) & \text{ if } n \in D_1 \text{ and } n \in D_2 \\ 1 & \text{ if } n \in D_1 \text{ but } n \notin D_2 \\ 1 & \text{ if } n \notin D_1 \text{ but } n \in D_2 \\ 0 & \text{ otherwise} \end{cases}
\]

Proposition 3.1. Let \( \langle x_n \rangle_{D_1} \) and \( \langle y_n \rangle_{D_2} \) be sequences over an outcome subspace \((\Omega, \leq, d, \hat{P})\). Let \( p \) be the outcome subspace sequence metric.

\( D_1 \cap D_2 \neq \emptyset \implies p \text{ is a metric} \)

Proof: This follows from the Fréchet product metric (Proposition D.9 page 159). In particular, \( p \) is a sum of metrics that include the metrics \( d(x_n, y_n) \) and the discrete metric (Definition D.8 page 155).

In standard signal processing, the autocorrelation of a sequence \( \langle x_n \rangle_{n \in D} \) is another sequence \( \langle y_n \rangle_{n \in D} \) defined as \( y_n \triangleq \sum_{m \in D} x_m x_{n-m} \). However, this definition requires that the sequence \( \langle x_n \rangle_{n \in D} \) be con-

---

1 quote: Aristotle (330BC?) (Book XIII Part 3)
image: http://upload.wikimedia.org/wikipedia/commons/9/98/Sanzio_01_Plato_Aristotle.jpg
structed over a field. In an outcome subspace sequence, we in general do not have a field; for example, in a die outcome subspace, the expressions $\boxplus + \boxminus$ and $\boxtimes \boxtimes$ are undefined. This paper offers an alternative definition (next) for autocorrelation that uses the distance $d$ and that does not require a field.

**Definition 3.2.** Let $(x_n)_{n \in \mathbb{N}}$ and $(y_n)_{n \in \mathbb{N}}$ be sequences over the outcome subspace $(\Omega, \leq, d, P)$. Let $p\left((x_n), (y_n)\right)$ be the outcome subspace sequence metric (Definition 3.1 page 77).

The cross-correlation $R_{xy}(n)$ of $(x_n)$ and $(y_n)$ and the autocorrelation $R_{xx}(n)$ of $(x_n)$ are defined as:

$$R_{xy}(n) \triangleq - \sum_{m \in \mathbb{Z}} p\left((x_{n+m}), (y_m)\right)$$

(cross-correlation $R_{xy}(n)$ of $(x_n)$ and $(y_n)$)

$$R_{xx}(n) \triangleq - \sum_{m \in \mathbb{Z}} p\left((x_{n+m}), (x_m)\right)$$

(autocorrelation $R_{xx}(n)$ of $(x_n)$)

Moreover, the $M$-offset autocorrelation of $(x_n)$ and $(y_n)$ is here defined as $R_{xx}(n) + M$ (Definition 1.44 page 24).

### 3.1.2 Examples of symbolic sequence statistics

**Example 3.1** (fair die sequence). Consider the pseudo-uniformly distributed fair die (Definition 2.7 page 28) sequence generated by the C code:

```c
#include<stdio.h>

int main()
{
    srand(0xSEED);
    for(n=0; n<N; n++) x[n] = 'A' + rand() % 6;
    return 0;
}
```

where ‘A’ represents $\bigcirc$, ‘B’ represents $\bigcirc$, ‘C’ represents $\bigcirc$, ‘D’ represents $\bigcirc$, ‘E’ represents $\bigcirc$.

The resulting sequence is partially displayed here:

```
0 1 2 3 4 5 6 7 8 9 10 11 12 13 ...
```

This sequence constrained to a length of $N = 2667 \times 6 = 16002$ elements is approximately uniformly distributed and uncorrelated, as illustrated next:

**Example 3.2** (real die sequence). Consider the pseudo-uniformly distributed real die (Definition 2.9 page 29) sequence generated as in Example 3.1 (page 78), but with the real die metric rather than the fair die metric. This change will not affect the distribution of the sequence, but it does affect the autocorrelation, as illustrated next:

**Example 3.3** (spinner sequence). Consider the pseudo-uniformly distributed spinner (Definition 2.10 page 29) sequence generated by the C code:

---

2For a more complete source code listing, see Section E2 (page 182)
#include <stdlib.h>

srand(0xSEED);

for (n=0; n<N; n++) \{ x[n] = \textquote{A'} + \texttt{rand()} \% 6; \}

The resulting sequence is partially displayed here:

This sequence is in essence identical to the fair die sequence (Example 3.1 page 78) and real die sequence (Example 3.2 page 78) and thus yields what is essentially an identical histogram. But because the metric is different, the autocorrelation is also different. In particular, because the nodes of the spinner metric are on average farther apart with respect to the spinner metric, the sequence is less correlated (with respect to the metric), as illustrated next:

Example 3.4 (weighted real die sequence). Consider the non-uniformly distributed \textit{weighted real die} (Definition 2.8 page 29) sequence with

\[
P(\text{\textbullet}) = 0.75 \text{ and } P(\text{\textcircled{1}}) = P(\text{\textcircled{2}}) = P(\text{\textcircled{3}}) = P(\text{\textcircled{4}}) = 0.05,
\]

generated by the C code\(^4\)

\[
\text{srand(0xSEED);}
\text{for (n=0; n<N; n++) \{ u = \texttt{rand()} \% 100;}
\text{if (\texttt{u<5}) x[n] = \textquote{A'}; /* 00-04 */
\text{else if (\texttt{u<10}) x[n] = \textquote{B'}; /* 05-09 */
\text{else if (\texttt{u<15}) x[n] = \textquote{C'}; /* 10-14 */
\text{else if (\texttt{u<20}) x[n] = \textquote{D'}; /* 15-19 */
\text{else if (\texttt{u<95}) x[n] = \textquote{E'}; /* 20-94 */
\text{else x[n] = \textquote{F'}; /* 95-99 */
\}
\]

The resulting sequence is partially displayed here:

Of course the resulting histogram, as illustrated below on the left, reflects the non-uniform distribution. Also note, as illustrated below on the right, that the weighted sequence is much more correlated (as defined by Definition 3.2 page 78) as compared to the uniformly distributed \textit{real die} sequence of Example 3.2 (page 78).

Example 3.5 (weighted die sequence). Consider the non-uniformly distributed \textit{weighted die} (Definition 2.6 page 28) sequence generated as in Example 3.4. Of course the resulting histogram is identical

\(^3\)For a more complete source code listing, see Section E.4 (page 193)

\(^4\)For a more complete source code listing, see Section E.3 (page 187)
to that of Example 3.4, but because the distance function is different, the autocorrelation sequence is also different.

Example 3.6 (weighted spinner sequence). Consider the non-uniformly distributed die sequence with

\[ P(5) = 0.75 \text{ and } P(1) = P(2) = P(3) = P(4) = P(6) = 0.05, \]

generated by the C code\(^5\)

```c
srand(0xSEED);
for(n=0; n<N; n++) { u = rand() %100;
  if (n >= 5) x[n] = 'A'; /* 00-04 */
  else if (n<10) x[n] = 'B'; /* 05-09 */
  else if (n<15) x[n] = 'C'; /* 10-14 */
  else if (n<20) x[n] = 'D'; /* 15-19 */
  else if (n<95) x[n] = 'E'; /* 20-94 */
  else x[n] = 'F'; /* 95-99 */
}
```

where 'A' represents ① , 'B' represents ② , 'C' represents ③ , 'D' represents ④ , 'E' represents ⑤ , and 'F' represents ⑥ .

The resulting sequence and histogram is in essence the same as in the weighted die sequence example (Example 3.4 page 79). But note, as illustrated below on the right, that the weighted spinner sequence of this example is significantly less correlated than the weighted die sequence of Example 3.4 (page 79), presumably due to the larger range (Definition 1.12 page 6) of the spinner metric \([0, 1, 2, 3]\) as compared to the range of the weighted real die metric \([0, 1, 2]\) and the weighted die metric \([0, 1]\).

Example 3.7 (Random DNA sequence). Consider the pseudo-uniformly distributed DNA sequence generated by the C code\(^6\)

```c
srand(0xSEED);
for(n=0; n<N; n++) { r = rand() %4;
  switch (r) { case 0: x[n] = 'A'; break;
      case 1: x[n] = 'T'; break;
      case 2: x[n] = 'C'; break;
      case 3: x[n] = 'G'; break; }
}
```

where 'A' represents A , 'T' represents T , 'C' represents C , and 'G' represents G .

This sequence constrained to a length of \(4000 \times 4 = 16000\) elements is approximately uniformly distributed and uncorrelated, as illustrated next:

Example 3.8 (SARS coronavirus DNA sequence). Consider the genome sequence (DNA sequence) for the SARS coronavirus with GenBank accession number NC_004718.3.\(^7\) This sequence is of

\(^5\)For a more complete source code listing, see Section E4 (page 193)

\(^6\)For a more complete source code listing, see Section E5 (page 199)

\(^7\)GenBank-NC_004718.3 (2011)
length $N = 29751$ and is partially displayed here, followed by its histogram and $2N$-offset auto-
correlation plots.

Example 3.9 (Ebola virus DNA sequence). Consider the genome sequence (DNA sequence) for the
Ebola virus with GenBank accession [AF086833.2]. This sequence is of length 18959 and is partially dis-
played here:

Example 3.10 (Bacterium DNA sequence). Consider the genome sequence (DNA sequence) for the
bacterium *Melissococcus plutonius strain 49.3 plasmid pMP19* with GenBank accession [NZ_CM003360.1].
This sequence is of length 19430 and is partially displayed here:

Example 3.11 (Papaya DNA sequence). Consider the genome sequence segment for the fruit *car-
ica papaya* with GenBank accession [DS982815.1]. This sequence is not a complete genome, rather it is
a “genomic scaffold” (Definition 2.12 page 29). As such, there are some elements for which the content is
not known. For these locations, the symbol ☐ is used. In this particular sequence, there are 144 ☐ symbols. This sequence is of length 15495 and is partially displayed here:

---

3.2 Extending to distance linear spaces

3.2.1 Motivation

Section 2.1 (page 27) demonstrated how a stochastic process could be defined as an outcome subspace with order and metric structures. Example 2.13 (page 53) reviewed an example of a real die outcome subspace that was mapped through 4 different random variables to 4 different weighted graphs. Two of these random variables (Y and Z) mapped to structures (weighted graphs) that are very similar to the real die with respect to order and metric geometry. Two other random variables (W and X) mapped to structures (the real line and the integer line) that are very dissimilar. The implication of this example is that if we want statistics that closely model the underlying stochastic process, then we should map to a structure that that an order structure and distance geometry similar to that of the underlying stochastic process, and not simply the one that is the most convenient. Ideally, we would like to map to a structure that is isomorphic (Definition 1.35 page 14) and isometric (Definition 1.37 page 20) to the structure of the stochastic process.

However, for sequence processing using very basic methods such as FIR filtering, Fourier analysis, or wavelet analysis, we would very much like to map into the real line $\mathbb{R}^1$ or possibly some higher dimensional space $\mathbb{R}^n$. Because the real line is often very dissimilar to the stochastic process, we are motivated to find structures in $\mathbb{R}^n$ that are similar. And that is what this section presents—mapping from a stochastic process $(\Omega, \leq, d, P)$ into an ordered distance linear space $(\mathbb{R}^n, \leq, d, +, \cdot, \mathbb{R}, +, \times)$ in which $\mathbb{R}^n$ is an extension of $\Omega$ and $d$ is an extension of $d$.

Thus, for sequence processing on an outcome subspace $(\Omega, \leq, d, P)$, we would like to define a random variable $X$ and an ordered distance linear space $(\mathbb{R}^n, \leq, d, +, \cdot, \mathbb{R}, +, \times)$ that satisfy the following constraints:

1. The random variable maps the elements of $\Omega$ into $\mathbb{R}^n$ and
2. the order relation $\leq$ is an extension to $\mathbb{R}^n$ of the order relation $\leq$ on $\Omega$ and
3. the distance function $d$ is an extension to $\mathbb{R}^n$ of the distance function $\hat{d}$ on $\Omega$.

3.2.2 Some random variables

In this section, we first define some random variables (Definition 2.13 page 46) that are used later in this paper.

**Definition 3.3.** The traditional die random variable $X$ maps from the set $\{\heartsuit, \clubsuit, \diamondsuit, \spadesuit, \clubsuit, \spadesuit\}$ into the set $\mathbb{R}^1$ and is defined as $^{11}$

$$X(\heartsuit) \triangleq 1, X(\clubsuit) \triangleq 2, X(\diamondsuit) \triangleq 3, X(\spadesuit) \triangleq 4, X(\clubsuit) \triangleq 5, \text{ and } X(\spadesuit) \triangleq 6.$$ 

**Definition 3.4.** The PAM die random variable $X$ maps from the set $\{\heartsuit, \clubsuit, \diamondsuit, \spadesuit, \clubsuit, \spadesuit\}$ into the set $\mathbb{R}^1$ and is defined as $^{12}$

$$X(\heartsuit) \triangleq -2.5, X(\clubsuit) \triangleq -1.5, X(\diamondsuit) \triangleq -0.5, X(\spadesuit) \triangleq +0.5, X(\clubsuit) \triangleq +1.5, \text{ and } X(\spadesuit) \triangleq +2.5.$$ 

**Version 0.50E**

A thesis concerning symbolic sequence processing

Daniel J. Greenhoe

2016 Jun 16 6:40AM UTC
Definition 3.5. The QPSK die random variable $X$ maps from the set $\{0, 1, 2, 3\}$ into the set $\mathbb{C}^2$ and is defined as:

\[
X(0) \triangleq \exp(30 \times \frac{\pi}{180}i), \quad X(1) \triangleq \exp(90 \times \frac{\pi}{180}i), \quad X(2) \triangleq \exp(150 \times \frac{\pi}{180}i), \\
X(3) \triangleq \exp(210 \times \frac{\pi}{180}i),
\]

Definition 3.6. The $\mathbb{R}^3$ die random variable $X$ maps from the set $\{0, 1, 2, 3, 4, 5\}$ into the set $\mathbb{R}^3$ and is defined as:

\[
X(0) \triangleq (0, 0, 0), \quad X(1) \triangleq (0, 1, 0), \quad X(2) \triangleq (1, 0, 0), \\
X(3) \triangleq (0, -1, 0), \quad X(4) \triangleq (1, -1, 0), \quad X(5) \triangleq (-1, 0, 0).
\]

Definition 3.7. The $\mathbb{R}^6$ die random variable $X$ maps from the set $\{0, 1, 2, 3, 4, 5\}$ into the set $\mathbb{R}^6$ and is defined as:

\[
X(0) \triangleq (0, 0, 0, 0, 0, 0), \quad X(1) \triangleq (0, 0, 1, 0, 0, 0), \\
X(2) \triangleq (0, 1, 0, 0, 0, 0), \quad X(3) \triangleq (0, 0, 0, 1, 0, 0), \quad X(4) \triangleq (0, 0, 0, 0, 1, 0), \quad X(5) \triangleq (0, 0, 0, 0, 0, 1).
\]

Definition 3.8. The $\mathbb{R}^1$ spinner random variable $X$ maps from the set $\{0, 1, 2, 3, 4, 5, 6\}$ into the set $\mathbb{R}$ and is defined as:

\[
X(0) \triangleq 1, \quad X(1) \triangleq 2, \quad X(2) \triangleq 3, \quad X(3) \triangleq 4, \quad X(4) \triangleq 5, \quad X(5) \triangleq 6.
\]

Definition 3.9. The QPSK spinner random variable $X$ maps from the set $\{0, 1, 2, 3, 4, 5, 6\}$ into the set $\mathbb{C}$ and is defined as:

\[
X(0) \triangleq \exp(-90 \times \frac{\pi}{180}i), \quad X(1) \triangleq \exp(-30 \times \frac{\pi}{180}i), \quad X(2) \triangleq \exp(30 \times \frac{\pi}{180}i), \\
X(3) \triangleq \exp(90 \times \frac{\pi}{180}i), \quad X(4) \triangleq \exp(150 \times \frac{\pi}{180}i), \quad X(5) \triangleq \exp(210 \times \frac{\pi}{180}i).
\]

Definition 3.10. The PAM DNA random variable $X$ maps from the set $\{0, 1, 2\}$ into the set $\mathbb{R}$ and is defined as:

\[
X(0) \triangleq -1.5, \quad X(1) \triangleq -0.5, \quad X(2) \triangleq +0.5.
\]

Definition 3.11. The QPSK DNA random variable $X$ maps from the set $\{0, 1, 2\}$ into the set $\mathbb{C}$ and is defined as:

\[
X(0) \triangleq \exp(45 \times \frac{\pi}{180}i), \quad X(1) \triangleq \exp(135 \times \frac{\pi}{180}i), \\
X(2) \triangleq \exp(225 \times \frac{\pi}{180}i), \quad X(3) \triangleq \exp(315 \times \frac{\pi}{180}i).
\]

Definition 3.12. The $\mathbb{R}^4$ DNA random variable $X$ maps from the set $\{0, 1, 2\}$ into the set $\mathbb{R}^4$ and is defined as:

\[
X(0) \triangleq (1, 0, 0, 0), \quad X(1) \triangleq (0, 1, 0, 0), \quad X(2) \triangleq (0, 0, 1, 0), \quad X(3) \triangleq (0, 0, 0, 1).
\]

3.2.3 Some ordered distance linear spaces

Definition 3.13. The structure $(\mathbb{R}, \leq, d)$ is the $\mathbb{R}^1$ die distance linear space if $\leq$ is the standard ordering relation on $\mathbb{R}$, and $d(x, y) \triangleq |x - y|$ (the Euclidean metric on $\mathbb{R}$, Definition D.10 page 156).

---

\(^{12}\)PAM is an acronym for pulse amplitude modulation and is a standard technique in the field of digital communications.

\(^{13}\)QPSK is an acronym for quadrature phase shift keying and is a standard technique in the field of digital communications.

\(^{14}\)Galleani and Garelo (2010) page 772

\(^{15}\)Galleani and Garelo (2010) page 772

\(^{16}\)QPSK is an acronym for quadrature phase shift keying and is a standard technique in the field of digital communications.

\(^{17}\)This type of mapping has previously been used by G. Voss (1992) in calculating the Voss Spectrum, (a kind of Fourier analysis) of DNA sequences. See also Galleani and Garelo (2010) page 772.
Definition 3.14. The structure \((\mathbb{R}^3, \leq, d)\) is the \(\mathbb{R}^3\) **die distance linear space** if \(\leq=\) \(\emptyset\), and \(d\) is the 2-scaled LAGRANGE ARC DISTANCE \(d\) defined as follows: \(d(p, q) \triangleq 2p(p, q)\) where \(p\) is the LAGRANGE ARC DISTANCE (Definition A.1 page 108).

![Diagram of die distance linear space](image1)

\[
\begin{array}{c|cccccccc}
  x & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
  y & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
  z & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
  \text{d(x, y)} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
  \end{array}
\]

Used together with the \(\mathbb{R}^3\) **die random variable** \(X\) (Definition 3.6 page 83), the distance \(d\) in the \(\mathbb{R}^3\) **die distance linear space** (Definition 3.14) is an extension of \(d\) in the **real die outcome subspace** \(G \triangleq (\{\Box, \Box, \Box, \Box, \Diamond, \Diamond\}, \leq, d, \emptyset)\) (Definition 2.9 page 29). We can also say that \(X\) is an **isometry** (Definition 1.37 page 20) and that the two structures are **isometric**. For example,
\[
d[X(\Box), X(\Box)] = d[(1, 0, 0), (0, 1, 0)] = 1 = d(\Box, \Box) \quad \text{and} \\
d[X(\Box), X(\Diamond)] = d[(1, 0, 0), (-1, 0, 0)] = 2 = d(\Box, \Box).
\]

As for order, the mapping \(X\) is also order preserving (Definition 1.34 page 14), but trivially, because the **real die outcome subspace** is unordered (Definition 1.24 page 10). But if we still honor the standard ordering on each dimension \(\mathbb{R}\) in \(\mathbb{R}^3\), then the two structures are **not isomorphic** (Definition 1.35 page 14) because\(^{19}\) the inverse \(X^{-1}\) is not order preserving (Theorem 1.5 page 14)—for example, \(X(\Box) = (0, 0, -1) \leq (0, 0, 1) = X(\Box)\), but \(\Box\) and \(\Box\) are incomparable (Definition 1.24 page 10) in \(G\).

Definition 3.15. The structure \((\mathbb{R}^2, \leq, d)\) is the \(\mathbb{R}^2\) **spinner distance linear space** if \(\leq=\) \(\emptyset\), and \(d\) is the 3-scaled LAGRANGE ARC DISTANCE \(d\) defined as follows: \(d(p, q) \triangleq 3p(p, q)\) where \(p\) is the LAGRANGE ARC DISTANCE (Definition A.1 page 108).

![Diagram of spinner distance linear space](image2)

\[
\begin{array}{c|cccccccc}
  x & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
  y & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
  \text{d(x, y)} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
  \end{array}
\]

Used together with the **QPSK spinner random variable** \(X\) (Definition 3.9 page 83), the distance \(d\) in the \(\mathbb{R}^2\) **spinner distance linear space** (Definition 3.15 page 84) is an extension of \(d\) in the **spinner outcome subspace** \(G \triangleq (\{\Box, \Box, \Box, \Box, \Diamond, \Diamond, \Diamond, \Diamond\}, \leq, d, \emptyset)\) (Definition 2.10 page 29). We can again say that \(X\) is an isometry and that the two structures are isometric. For example,
\[
d[X(\Box), X(\Box)] = d[(0, -1), (\sqrt{2}, -\sqrt{2})] = 1 = d(\Box, \Box) \quad \text{and} \\
d[X(\Box), X(\Diamond)] = d[(0, -1), (\sqrt{2}, +\sqrt{2})] = 2 = d(\Box, \Box) \quad \text{and} \\
d[X(\Box), X(\Diamond)] = d[(0, -1), (0, 1)] = 3 = d(\Box, \Box).
\]

The mapping \(X\) is again trivially order preserving. And if we again honor the standard ordering on each dimension \(\mathbb{R}\) in \(\mathbb{R}^2\), then the two structures are **not isomorphic** (Definition 1.35 page 14) because the inverse \(X^{-1}\) is not order preserving—for example, \(X(\Box) = (0, -1) \leq (0, 1) = X(\Box)\), but \(\Box\) and \(\Box\) are incomparable in \(G\).

Definition 3.16. The structure \((\mathbb{R}^6, \leq, d)\) is the \(\mathbb{R}^6\) **die distance linear space** if \(\leq=\) \(\emptyset\), and \(d\) is defined as \(d(p, q) \triangleq \sqrt{2}p(p, q)\), where \(p\) is the **EUCLIDEAN METRIC** on \(\mathbb{R}^6\) (Definition D.10 page 156).

![Diagram of die distance linear space](image3)

\[
\begin{array}{c|cccccccc}
  x & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
  y & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
  z & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
  w & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\
  \text{d(x, y)} & 0 & 1 & 2 & 3 & 4 & 5 & 6 & 7 \\
  \end{array}
\]

Used together with the \(\mathbb{R}^6\) **die random variable** \(X\) (Definition 3.7 page 83), the distance \(d\) in the \(\mathbb{R}^6\) **fair die distance linear space** (Definition 3.16 page 84) is an extension of \(d\) in the **fair die outcome subspace**

\(^{19}\)Note that while \(X^{-1}\) (Definition 1.7 page 6) does not exist as a function, it does exist as a relation.
\[
G \triangleq \{ \square, \circ, \Diamond, \triangle, \diamond \}, \preceq, \hat{d}, \hat{\hat{d}} \text{ (Definition 2.9 page 29).}
\]
We can again say that \( X \) is an isometry and that the two structures are isometric. For example,
\[
d(X(\square), X(\square)) = d((1, 0, 0, 0, 0), (0, 1, 0, 0, 0)) = 1 = \hat{d}(\square, \square) \text{ and }
\]
\[
d(X(\square), X(\square)) = d((1, 0, 0, 0, 0), (0, 0, 1, 0, 0)) = 1 = \hat{d}(\square, \diamond) \text{ and }
\]
\[
d(X(\square), X(\square)) = d((1, 0, 0, 0, 0), (0, 0, 0, 0, 1)) = 1 = \hat{d}(\square, \Box).
\]
The mapping \( X \) is again trivially order preserving, and the inverse \( X^{-1} \) is trivially order preserving as well. And so unlike the \( \mathbb{R}^3 \) die distance linear space (Definition 3.14) and the \( \mathbb{R}^2 \) spinner distance linear space (Definition 3.15), this pair of structures is isomorphic.

### 3.3 Symbolic sequence processing applications

**“I regard as quite useless the reading of large treatises of pure analysis: too large a number of methods pass at once before the eyes. It is in the works of applications that one must study them; one judges their ability there and one apprises the manner of making use of them.”**

Joseph Louis Lagrange (1736-1813), mathematician

#### 3.3.1 Low pass filtering/Smoothing

*Example 3.12* (low pass filtering of real die sequence).

1. Consider the pseudo-uniformly distributed die sequence presented in Example 3.2 (page 78). Suppose we want to filter this sequence with a low pass sequence in order to “smooth out” the sequence. But to perform the actual filtering, note that the die sequence must first be mapped into a linear space \( \mathbb{R}^N \).

2. Suppose we first use the traditional die random variable (Definition 3.3 page 82) to map the die sequence into \( \mathbb{R}^1 \). Filtering (Definition 1.45 page 25) this \( \mathbb{R} \)-valued sequence using the length 16 rectangular low pass sequence (Example 1.18 page 25) in the \( \mathbb{R}^1 \) die distance linear space (Definition 3.13 page 83) and then mapping the result back to a die sequence using the Euclidean metric (Definition D.10 page 156), produces the result partially displayed here:

![Die sequence diagram](http://en.wikipedia.org/wiki/Image:Langrange_portrait.jpg)

Note that the die sequence has indeed been smoothed out, but its uniform distribution has been destroyed—almost all of its values are around the “expected value” 3.5, as illustrated below on the left. Of course such filtering also introduces correlation, giving the autocorrelation sequence a slightly wider center lobe as illustrated below on the right. Both diagrams are

---

\[^{20}\text{quote: } \text{Stopple (2003), page xi} \]

\[^{20}\text{image: } \text{http://en.wikipedia.org/wiki/Image:Langrange_portrait.jpg} \]
3. Alternatively, suppose we next try using the \( \mathbb{R}^3 \) die random variable (Definition 3.6 page 83) to map the die sequence into \( \mathbb{R}^3 \). Filtering this new sequence using the length 16 rectangular low pass sequence in the \( \mathbb{R}^3 \) distance linear space (Definition 3.14 page 84) and then mapping back to a die sequence using the Lagrange arc distance yields the result partially displayed here:

Note that the die sequence does appear to be “smoothed out”, but this time the distribution is much more uniform, as illustrated below on the left; and is slightly less correlated (12795 compared to 17658), as illustrated below on the right.

4. Using a length 16 Hanning low pass sequence (Definition 1.48 page 25) rather than the length 16 rectangular low pass sequence as in item (3) results in a distribution that is more uniform and in a sequence that is very slightly less correlated:

5. Using a length 50 Hanning low pass sequence (Example 1.19 page 25) rather than the length 16 Hanning low pass sequence as in item (4) results in about the same uniformity of distribution, about 1.8% lower side lobes in the autocorrelation sequence \( \frac{12733 - 12505}{12733} \times 100 \approx 1.8 \), but a wider main lobe (presumably due to the longer filter width):

6. Using a length 50 rectangular low pass sequence rather than the length 50 Hanning low pass sequence as in item (5) results in a distribution that is a little less uniform and about 3.3%
more correlated ($\frac{1205\text{ to }1206}{1205} \times 100 \approx 3.3$):

\[
0 \quad 2052 \quad 1895 \quad 1894 \quad 1944 \quad 1927 \quad 2378
\]

(16.96%) (15.66%) (14.91%) (16.07%) (15.93%) (19.65%)

7. Replacing the Lagrange arc distance by the Euclidean metric in this example has very little effect. More details follow:

(a) Using the Euclidean metric in $\mathbb{R}^3$ rather than the Lagrange arc distance in item (3) yields sequences that are identical.\(^{21}\)

(b) Using the Euclidean metric in item (4) rather than the Lagrange arc distance yields sequences that differ at 6 locations out of $N + M + M - 1 = 12000 + 16 + 16 - 1 = 12031$ locations (differ at approximately 0.05% of the locations).\(^{22}\)

<table>
<thead>
<tr>
<th>n</th>
<th>Euclidean</th>
<th>Lagrange</th>
</tr>
</thead>
<tbody>
<tr>
<td>281</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1630</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11888</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(c) Using the Euclidean metric in $\mathbb{R}^3$ rather than the Lagrange arc distance as in item (5) (length 50 Hanning filter) yields sequences that are identical.\(^ {23}\)

(d) Using the Euclidean metric in $\mathbb{R}^3$ rather than the Lagrange arc distance as in item (6) (length 50 rectangular filter) differ at 85 locations out of $N + M + M - 1 = 12000 + 50 + 50 - 1 = 12099$ locations (differ at approximately 0.7% of the locations).\(^ {24}\)

**Example 3.13** (low pass filtering of spinner sequence).

1. Consider the pseudo-uniformly distributed spinner sequence presented in Example 3.3 (page 78). As in Example 3.12 (page 85), suppose we want to filter this sequence with a low pass rectangular sequence in order to “smooth out” the sequence.

2. Suppose we first use the $\mathbb{R}^1$ spinner random variable (Definition 3.8 page 83) to map the spinner sequence into $\mathbb{R}^1$. Filtering this mapped sequence using the length 16 rectangular low pass sequence and then mapping the result back to a spinner sequence using the Euclidean metric, produces the result partially displayed here (in essence the same as in Example 3.12 page 85):

Again, it's uniform distribution has been essentially destroyed.

\[^{21}\text{See experiment log file "rdie_lp_12000m16.xlg" generated by the program "ssp.exe".}\]

\[^{22}\text{See experiment log file "rdie_lp_12000m16.xlg" generated by the program "ssp.exe".}\]

\[^{23}\text{See experiment log file "rdie_lp_12000m50.xlg" generated by the program "ssp.exe".}\]

\[^{24}\text{See experiment log file "rdie_lp_12000m50.xlg" generated by the program "ssp.exe".}\]
3. Alternatively, suppose we next try using the QPSK spinner random variable (Definition 3.9 page 83) to map the spinner sequence into $\mathbb{C} \cong \mathbb{R}^2$. Filtering this new sequence using the length 16 rectangular low pass sequence in the $\mathbb{R}^2$ spinner distance linear space (Definition 3.15 page 84) and then mapping back to a sequence over the spinner outcome subspace using the Lagrange arc distance yields the result partially displayed here:

Note that the sequence does appear to be “smoothed out”, but this time the distribution is much more uniform and about 69% less correlated than the $\mathbb{R}^1$ method of item (2):

Furthermore, it is about 58% less correlated than the $\mathbb{R}^3$ filtering for the die sequence used in item (3) of Example 3.12 (page 85).

4. Using the Euclidean metric rather than the Lagrange arc distance as in item (3) results in a sequence that differs at 99 different locations out of $N + M + M - 1 = 12000 + 16 + 15 = 12031$ locations (approximately 0.8% of the locations differ).\(^\text{25}\)

5. Using a length 16 Hanning low pass sequence rather than the length 16 Rectangular low pass sequence as in item (3) results in a distribution that is more uniform and about 5.3% less correlated:

6. Using the Euclidean metric rather than the Lagrange arc distance as in item (5) results in a sequence that differs at exactly 2 locations (approximately 0.017%) out of 12031 locations.\(^\text{26}\)

7. Using a length 50 Hanning low pass sequence rather than the length 16 Hanning low pass sequence as in item (5) results in the following:

\(^{25}\)See experiment log file “spin_lp_12000m16.xlg” generated by the program “ssp.exe”.

\(^{26}\)See experiment log file “spin_lp_12000m16.xlg” generated by the program “ssp.exe”.
8. Using a length 50 Rectangular low pass sequence rather than the length 50 Hanning low pass sequence as in item (7) results in the following:

Example 3.14 (low pass filtering of fair die sequence).

1. Consider the pseudo-uniformly distributed die sequence presented in Example 3.1 (page 78). Suppose we want to filter this sequence with a low pass sequence in order to “smooth out” the sequence, just as in Example 3.12 (page 85).

2. Suppose we first use the traditional die random variable (Definition 3.3 page 82) to map the die sequence into \( \mathbb{R}^1 \). Filtering this mapped sequence using the length 16 rectangular low pass sequence and then mapping the result back to a die sequence using the Euclidean metric, produces a result identical to that of item (2) (page 85) of Example 3.12.

3. Alternatively, suppose we next use the \( \mathbb{R}^6 \) die random variable (Definition 3.7 page 83) to map the die sequence into \( \mathbb{R}^6 \). Filtering this new sequence using the length 16 rectangular low pass sequence in the \( \mathbb{R}^6 \) die distance linear space (Definition 3.16 page 84) and then mapping back to a die sequence using the Euclidean metric yields a much more uniform distribution and a sequence that is about 28% less correlated.

![Image showing a graph of a low pass sequence](image)

Note further that this \( \mathbb{R}^6 \) technique yeilds a sequence that is about 9.9% more correlated than yielded by the \( \mathbb{R}^3 \) technique used in item (3) of Example 3.12 (page 85).

4. Using a length 16 Hanning low pass sequence rather than the length 16 Rectangular low pass sequence as in item (3) results in a distribution that is more uniform and a sequence that is about 0.12% less correlated:

![Image showing a graph of a low pass sequence](image)
5. Using a length 50 Hanning low pass sequence rather than the length 16 Hanning low pass sequence as in item (4) results in the following:

Note further that this \( \mathbb{R}^6 \) technique yields a sequence that is about 11% more correlated than yielded by the \( \mathbb{R}^3 \) technique used in item (5) of Example 3.12 (page 85).

6. Using a length 50 Rectangular low pass sequence rather than the length 50 Hanning low pass sequence as in item (5) results in the following:

Note further that this \( \mathbb{R}^6 \) technique yields a sequence that is about 8.8% more correlated than yielded by the \( \mathbb{R}^3 \) technique used in item (6) of Example 3.12 (page 85).

7. In the fair die outcome space, the Lagrange arc distance does not seem so appropriate. That being said however, ...

(a) using the Lagrange arc distance rather than the Euclidean metric in item (3) page 89 yields results that are identical.

(b) using the Lagrange arc distance rather than the Euclidean metric in item (4) page 89 yields results that differ at 5 locations (differ at approximately 0.04% of the total possible \( N + M + M - 1 = 12000 + 16 + 16 - 1 = 12031 \) locations).

(c) using the Lagrange arc distance rather than the Euclidean metric in item (5) page 90 yields results that are identical.

(d) using the Lagrange arc distance rather than the Euclidean metric in item (6) page 90 yields results that differ at 289 locations (differ at approximately 2.4% of the total possible 12031 locations).

8. Empirical evidence observed in items 3, 4, 5, and 6, suggests that the \( \mathbb{R}^6 \) technique of this example leads to about 10% more correlation than the \( \mathbb{R}^3 \) technique of Example 3.12 (page 85).

---

27 See experiment log file “idle_lp_12000m16.xlg” generated by the program “ssp.exe”.

28 See experiment log file “idle_lp_12000m16.xlg” generated by the program “ssp.exe”.

29 See experiment log file “idle_lp_12000m50.xlg” generated by the program “ssp.exe”.

30 See experiment log file “idle_lp_12000m50.xlg” generated by the program “ssp.exe”.
3.3.2 High pass filtering

Example 3.15 (high pass filtering of weighted real die sequence).

1. Consider a length \(50(1200 + 2) - (50 - 1) = 60051\) non-uniformly distributed die sequence generated as described in Example 3.4 (page 79). To remove the strong bias, we could map and filter (Definition 1.45 page 25) the sequence with the length 50 high pass rectangular sequence (Definition 1.47 page 25). Such filtering will obviously introduce correlation into the die sequence. The low pass filtering of Example 3.12 page 85 (“smoothing”) also introduced correlation, but wanting a “smooth” sequence informally implies a willingness to accept a highly correlated sequence. However in this current example, we would prefer to have an uncorrelated sequence. To negate the correlation introduced by filtering, we down sample (Definition 1.42 page 23) the filtered sequence by a factor of 50 and remove the first and last element, leaving a sequence of length 1200.

2. If the filtering and downsampling described in item (1) is performed in the traditional \(\mathbb{R}^1\) space, then after mapping back to a die sequence using the Euclidean metric, we obtain the result partially displayed here...

3. Alternatively, suppose we next use the \(\mathbb{R}^3\) die random variable (Definition 3.6 page 83) to map the die sequence into \(\mathbb{R}^3\). Filtering this new sequence using the length 50 rectangular high pass sequence in the \(\mathbb{R}^3\) distance linear space (Definition 3.14 page 84) and then mapping back to a die sequence using the Lagrange arc distance yields the following results:

Note that neither the \(\mathbb{R}^1\) method of item (2) nor the \(\mathbb{R}^3\) method of item (3) yields a uniformly distributed sequence; but the \(\mathbb{R}^3\) method at least comes significantly closer to this end. Moreover, the \(\mathbb{R}^3\) method also yields a sequence that is less correlated.

4. Replacing the length 50 rectangular high pass filter in item (3) with the length 50 Hanning high pass filter (Definition 1.49 page 25) yields a different sequence with similar distribution but is
5. Replacing the Lagrange arc distance by the Euclidean metric in this example has very little effect, even before downsampling. Before downsampling, the length of each sequence is \( M(N + 2) = 50(1202) = 60100 \) elements. More details follow:

(a) Using the Euclidean metric rather than the Lagrange arc distance in item (3) yields results that are identical.\(^\text{31}\)

(b) Using the Euclidean metric rather than the Lagrange arc distance in item (4) yields results that differ at 4 locations (approximately 0.007% of all the locations).\(^\text{32}\)

6. For the type of sequence processing described in this example, item (5) very informally suggests the following:

(a) The processing is not highly sensitive to the choice of distance function.

(b) The processing is not heavily dependent on the triangle inequality.

Example 3.16 (high pass filtering of weighted spinner sequence).

1. Consider a length 50(1200 + 2) – (50 – 1) = 60051 non-uniformly distributed spinner sequence generated as described in Example 3.6 (page 80). To remove the strong \( \odot \) bias, we could filter the sequence with the length 50 high pass rectangular sequence and down sample the filtered sequence by a factor of 50, as described in Example 3.15 (page 91).

2. If the filtering described in item (1) is performed in the traditional \( \mathbb{R}^1 \) space, then after mapping back to a spinner sequence using the Euclidean metric, we obtain the result partially displayed here...

where the bias at \( \odot \) has been replaced by a new bias at \( \odot \), as illustrated quantitatively below on the left, calculated over 1200 elements.

\(^\text{31}\) See experiment log file “wdie_hp_1200m50.xlg” generated by the program “ssp.exe”.

\(^\text{32}\) See experiment log file “wdie_hp_1200m50.xlg” generated by the program “ssp.exe”.
3. If we replace the *length 50 rectangular high pass filter* of item (2) with a *length 50 Hanning high pass filter* then we obtain the result partially displayed here:

![Diagram showing the rectangular high pass filter](image)

...where the bias at ⑤ again has been replaced by a new bias at ①:

![Diagram showing the Hanning high pass filter](image)

4. If the rectangular filtering in \( \mathbb{R}^1 \) of item (2) is instead performed in \( \mathbb{R}^2 \) and mapped back to a *spinner sequence* using the *Lagrange arc distance*, then we obtain the result partially displayed here:

![Diagram showing the rectangular filtering in \( \mathbb{R}^2 \)](image)

Note that neither the \( \mathbb{R}^1 \) methods (described in item (2) and item (3)) nor the \( \mathbb{R}^2 \) method (described in item (4)) yields a uniformly distributed sequence; but the \( \mathbb{R}^2 \) method at least comes significantly closer to this end.

5. Replacing the *Lagrange arc distance* by the *Euclidean metric* as in item (4) yields a sequence that differs at a total of 272 locations (approximately 0.5% of the locations).\(^{33}\)

6. If instead of using the rectangular filtering (as in item (4)), we use the Hanning filtering of item (3) in \( \mathbb{R}^2 \) and map back to a *spinner sequence* using the *Lagrange arc distance*, then we obtain the result partially displayed here:

![Diagram showing the Hanning filtering in \( \mathbb{R}^2 \)](image)

7. Replacing the *Lagrange arc distance* by the *Euclidean metric* in item (6) yields a sequence that differs at a total of 3 locations (approximately 0.005% of the locations).\(^{34}\)

---

\(^{33}\) See experiment log file “wsin_hp_1200m50.xlg” generated by the program “ssp.exe”.

\(^{34}\) See experiment log file “wsin_hp_1200m50.xlg” generated by the program “ssp.exe”.
Example 3.17 (high pass filtering of weighted die sequence).

1. Consider a length 50(1200 + 2) − (50 − 1) = 60051 weighted die sequence generated as described in Example 3.5 (page 79). To remove the strong □ biais, we could map and filter the sequence with the length 16 high pass rectangular sequence (Example 1.18 page 25). To negate the correlation introduced by filtering, we down sample the filtered sequence by a factor of 16.

2. If the die sequence of item (1) is mapped into \( \mathbb{R}^1 \) using the traditional die random variable (Definition 3.3 page 82), filtered, down sampled, and mapped back to a die sequence using the Euclidean metric, we obtain the result partially displayed here...

   ![Graph of data sequence](image)

   where the bias at □ has been replaced by a new bias at □, as illustrated quantitatively below on the left, calculated over 1200 elements.

3. But if instead of processing the die sequence in \( \mathbb{R}^1 \) as in item (2), processing is performed in \( \mathbb{R}^6 \) and mapped back to a die sequence using the Euclidean metric, then we obtain the result partially displayed here:

   ![Graph of data sequence](image)

4. Replacing the length 16 rectangular sequence in item (3) with a length 16 Hanning sequence in \( \mathbb{R}^6 \) yields the following results:

   ![Graph of data sequence](image)
5. Replacing the length 16 Hanning sequence in item (4) with a length 50 Hanning sequence in \( \mathbb{R}^6 \) yields the following results:

![Graph showing a comparison between two sequences.]

Note that this \( \mathbb{R}^6 \) technique yields a sequence that is about 8.7% more correlated than yielded by the \( \mathbb{R}^3 \) technique used in item (4) of Example 3.15 (page 91).

6. Replacing the length 50 Hanning sequence in item (5) with a length 50 rectangular sequence in \( \mathbb{R}^6 \) yields the following results:

![Graph showing a comparison between two sequences.]

Note that this \( \mathbb{R}^6 \) technique yields a sequence that is about 7.3% more correlated than yielded by the \( \mathbb{R}^3 \) technique used in item (3) of Example 3.15 (page 91).

7. As in Example 3.14 (page 89), here again the Lagrange arc distance does not seem so appropriate. That again being said however, ...

(a) using the Lagrange arc distance rather than the Euclidean metric in item (3) page 94 yields results that are identical. \(^{35}\)

(b) using the Lagrange arc distance rather than the Euclidean metric in item (4) page 94 yields results that differ at 17 locations (differ at approximately 0.09% of the total possible \( M(N+2) = 16(1200 + 2) = 19232 \) locations). \(^{36}\)

8. Empirical evidence observed in items item (5) and item (6) suggests that the \( \mathbb{R}^6 \) technique of this example leads to about 8% more correlation than the \( \mathbb{R}^3 \) technique of Example 3.15 (page 91).

### 3.3.3 Fourier Analysis

**Example 3.18** (length 1200 non-stationary die sequence with 10Hz oscillating mean).

---

\(^{35}\) See experiment log file “wdie_hp_1200m16.xlg” generated by the program “ssp.exe”.

\(^{36}\) See experiment log file “wdie_hp_1200m16.xlg” generated by the program “ssp.exe”.

Daniel J. Greenhoe
1. Suppose we have a length $N \triangleq 1200$ die sequence $(x_n)$ with the following distribution:
\[
P(\square) = P(\square) = P(\square) = P(\square) = 0.15 \quad \text{and} \quad P(\square) = 0.25
\]
for $n \in \{p + (2m)\frac{N}{M} | p = 0, 1, \ldots, \frac{N}{M} - 1, m = 0, 1, 2, \ldots, 9\}$ and
\[
P(\square) = P(\square) = P(\square) = P(\square) = 0.15 \quad \text{and} \quad P(\square) = 0.25
\]
for $n \in \{p + (2m + 1)\frac{N}{M} | p = 0, 1, \ldots, \frac{N}{M} - 1, m = 0, 1, 2, \ldots, 9\}$
where $M \triangleq 120$. That is, the distribution of the sequence oscillates every $\frac{M}{2} = 60$ samples between one that favors $\square$ and one that favors $\square$. Moreover, if we were to evaluate the sequence using a Discrete Fourier Transform operator $\mathcal{F}$ (Definition 1.52 page 26), we might expect to see a strong component at $\frac{N}{M} = 10$ (or 10 Hz—the distribution goes through 10 cycles during the course of the sequence).

2. Suppose we first use the PAM die random variable (Definition 3.4 page 82) to map the sequence of item (1) into $\mathbb{R}^1$. The magnitude of the $\mathcal{F} : \mathbb{R}^1 \rightarrow \mathbb{C}^1$ of the mapped sequence is as follows:

Looking at the above result, it would be next to impossible to discern that the distribution had a significantly strong oscillation of 10 cycles. In fact, the magnitude of the DFT at 10Hz is only 0.895699, or $10 \log_{10}(0.895699) = -0.478377$ dB. There are exactly 456 out of a total $\frac{N}{2} = 600$ values that are greater than the DFT magnitude at 10Hz.\(^{37}\) That is, to either a human observer or a machine algorithm, the 10Hz component is effectively lost in the noise.

3. Suppose we next use the QPSK die random variable (Definition 3.5 page 83) to map the sequence into the complex plane. The magnitude of the $\mathcal{F} : \mathbb{C}^1 \rightarrow \mathbb{C}^1$ operation on the mapped sequence is as follows:

The magnitude of the DFT at 10Hz is 0.589990, or $10 \log_{10}(0.589990) = -2.291552$ dB. There are exactly 831 out of a total $N = 1200$ values that are greater than the DFT magnitude at 10Hz.\(^{38}\) Again, the 10Hz component is effectively lost in the noise.

4. Suppose we next use the $\mathbb{R}^6$ die random variable (Definition 3.7 page 83) to map the sequence into $\mathbb{R}^6$. The magnitude of $\mathcal{F} : \mathbb{R}^6 \rightarrow \mathbb{C}^6$ of the mapped sequence is as follows:

The magnitude at 10Hz is 1.556295, or $10 \log_{10}(1.556295) = 1.920920$ dB. Besides the DC component (0Hz component), this is the uniquely greatest value of the 600 samples. And in fact,

\(^{37}\) See experiment log file “diedft_1525_1200m120_xlq” generated by the program “ssp.exe”.

\(^{38}\) See experiment log file “diedft_1525_1200m120_xlq” generated by the program “ssp.exe”. 
there are only 5 out of a total $\frac{N}{M} = 600$ samples that are $0.90 \times 1.556295$ or greater. Therefore, using the $\mathbb{R}^6$ mapping technique of this example, it is much simpler to detect the 10Hz oscillating distribution.

**Example 3.19** (length 12000 non-stationary die sequence with 10Hz oscillating mean).

1. Suppose we have a length $N = 12000$ die sequence $\{x_n\}$ with the following distribution:

$$P(\square) = P(\bigcirc) = P(\blacksquare) = P(\bullet) = P(\star) = 0.16$$

$$P(\triangle) = P(\blacktriangle) = 0.20$$

for $n \in \{ p + (2m) \frac{M}{N} | p = 0, 1, \ldots, \frac{M}{2} - 1, m = 0, 1, 2, \ldots, 9 \}$ and

$$P(\square) = P(\bigcirc) = P(\blacksquare) = P(\bullet) = P(\star) = 0.16$$

$$P(\triangle) = P(\blacktriangle) = 0.20$$

for $n \in \{ p + (2m + 1) \frac{M}{N} | p = 0, 1, \ldots, \frac{M}{2} - 1, m = 0, 1, 2, \ldots, 9 \}$

where $M = 1200$. That is, the distribution of the sequence oscillates every $\frac{M}{M} = 600$ samples between one that favors $\square$ and one that favors $\Delta$. If we were to evaluate the distribution using the *Discrete Fourier Transform* operator, we again might expect to see a strong component at $\frac{N}{M} = 10$ (or 10 Hz—the distribution goes through 10 cycles during the course of the sequence).

2. Suppose we first use the *PAM die random variable* (Definition 3.4 page 82) to map the sequence of item (1) into $\mathbb{R}^1$. In the magnitude of $\mathcal{DFT} : \mathbb{R}^1 \to \mathbb{C}^1$ there are 1130 values out of a possible $\frac{N}{M} = 6000$ values greater than the value at 10Hz (that value being 2.174512). As in Example 3.18 (page 95), the subtle 10Hz component is effectively lost in the noise.

3. Suppose we next use the *QPSK die random variable* (Definition 3.5 page 83) to map the sequence into the complex plane. There are exactly 1932 out of a total $N = 12000$ values that are greater than the DFT value at 10Hz (that value being 1.348693). As in Example 3.18 (page 95), the subtle 10Hz component is effectively lost in the noise.

4. Suppose we next use the $\mathbb{R}^6$ *die random variable* (Definition 3.7 page 83) to map the sequence into $\mathbb{R}^6$. The magnitude of $\mathcal{DFT} : \mathbb{R}^6 \to \mathbb{C}^6$ of the mapped sequence is as follows:

![Magnitude of DFT](chart.png)

Besides the DC component, the value at 100Hz (that value being 1.965018) is the uniquely greatest value of the $\frac{N}{M} = 6000$ samples; and it is $10 \log_{10}(1.965018/1.660189) = 0.699 \ldots$ dB larger than the next largest value. Thus, even though the oscillating distribution is very subtle (even more subtle than that of Example 3.18 (page 95)), the $\mathbb{R}^6$ mapping technique and subsequent analysis are able to detect it.

**Example 3.20** (length 12000 non-stationary die sequence with 10Hz oscillating mean).

---

39 See experiment log file “diedft_1525_1200m120.xvg” generated by the program “ssp.exe”. The 5 largest values are the points (0, 14.251433), (10, 1.556295), (344, 1.513501), (456, 1.405843) and (557, 1.468970).

40 See experiment log file “diedft_1620_1200m120.xvg” generated by the program “ssp.exe”.

41 See experiment log file “diedft_1620_1200m120.xvg” generated by the program “ssp.exe”.

42 See experiment log file “diedft_1620_1200m120.xvg” generated by the program “ssp.exe”. The 10 largest values are (0, 44.763194), (10, 1.965018), (90, 1.602474), (1223, 1.660189), (1313, 1.553349), (2385, 1.551028), (3039, 1.550918), (4154, 1.563756), (4187, 1.586362), and (5147, 1.623052).
1. Suppose we have a length $N \triangleq 12000$ die sequence $(x_n)$ with the following distribution:

$$P(\Box) = P(\mathbb{R}) = P(\oplus) = P(\otimes) = 0.16 \quad \text{and} \quad P(\circ) = 0.20$$

for $n \in \{ p + (2m) M_p | p = 0, 1, \ldots, \frac{M_p}{2} - 1, m = 0, 1, 2, \ldots, 9 \}$ \hspace{1cm} \text{and} \hspace{1cm}

$$P(\Box) = P(\mathbb{R}) = P(\oplus) = P(\otimes) = 0.16 \quad \text{and} \quad P(\circ) = 0.20$$

for $n \in \{ p + (2m + 1) M_p | p = 0, 1, \ldots, \frac{M_p}{2} - 1, m = 0, 1, 2, \ldots, 9 \}$

where $M \triangleq 120$. That is, the distribution of the sequence oscillates every $\frac{M_p}{2} = 600$ samples between one that favors $\Box$ and one that favors $\mathbb{R}$. If we were to evaluate the sequence using the Discrete Fourier Transform operator, we might expect to see a strong component at $\frac{N}{M} = 100$ (or 100 Hz—the distribution goes through 100 cycles during the course of the sequence).

2. Suppose we first use the PAM die random variable (Definition 3.4 page 82) to map the sequence of item (1) into $\mathbb{R}^6$. In the magnitude $\text{DFT} : \mathbb{R}^1 \to \mathbb{C}^1$ there are 1320 values out of a possible $\frac{N}{2} = 6000$ values greater than the value at 100Hz that value being 2.081469. The subtle 100Hz component is effectively lost in the noise.

3. Suppose we next use the QPSK die random variable (Definition 3.5 page 83) to map the sequence into the complex plane. There are exactly 1555 out of a total $N12000$ values that are greater than the DFT value at 100Hz that value being 1.425427. The subtle 100Hz component is effectively lost in the noise.

4. Suppose we next use the $\mathbb{R}^6$ die random variable (Definition 3.7 page 83) to map the sequence into $\mathbb{R}^6$. The magnitude of $\text{DFT} : \mathbb{R}^6 \to \mathbb{C}^6$ of the mapped sequence is as follows:

![DFT Magnitude Plot](image)

Besides the DC component, the value at 100Hz (that value being 2.256927) is the uniquely greatest value of the $\frac{N}{2} = 600$. and it is $10 \log_{10}(2.256927/1.599335) = 1.495 \quad \text{dB}$ larger than the next largest value. Thus, even though the oscillating distribution is very subtle, the $\mathbb{R}^6$ mapping technique and subsequent analysis are able to detect it.

**Example 3.21** (length 12000 non-stationary artificial DNA sequence with 10Hz oscillating mean).

1. Suppose we have a length $N \triangleq 12000$ die sequence $(x_n)$ with the following distribution:

$$P(\Box) = P(\mathbb{R}) = P(\oplus) = P(\otimes) = 0.24 \quad \text{and} \quad P(\circ) = 0.28$$

for $n \in \{ p + 2m M_p | p = 0, 1, \ldots, \frac{M_p}{2} - 1, m = 0, 1, 2, \ldots, 9 \}$ \hspace{1cm} \text{and} \hspace{1cm}

$$P(\Box) = P(\mathbb{R}) = P(\oplus) = P(\otimes) = 0.24 \quad \text{and} \quad P(\circ) = 0.28$$

for $n \in \{ p + (2m + 1) M_p | p = 0, 1, \ldots, \frac{M_p}{2} - 1, m = 0, 1, 2, \ldots, 9 \}$

where $M \triangleq 1200$. That is, the distribution of the sequence oscillates every $\frac{M_p}{2} = 600$ samples between one that favors $\Box$ and one that favors $\mathbb{R}$. Moreover, if we were to evaluate the sequence using a Discrete Fourier Transform (DFT) operator, we might expect to see a strong component at $\frac{N}{M} = 10$ (or 10 Hz—the distribution goes through 10 cycles during the course of the sequence).

---

43 See experiment log file “diedft_1620_12000m120.xlg” generated by the program “ssp.exe”.

44 See experiment log file “diedft_1620_12000m120.xlg” generated by the program “ssp.exe”.

45 See experiment log file “diedft_1620_12000m120.xlg” generated by the program “ssp.exe”. The 10 largest values are (0, 44.763060), (90, 1.577597), (100, 2.256927), (486, 1.599335), (1223, 1.585154), (1313, 1.547956), (3039, 1.553522), (3162, 1.561863), (5147, 1.558487), and (5567, 1.533659).
2. Suppose we first use the **PAM DNA random variable** (Definition 3.10 page 83) to map the DNA sequence into $\mathbb{R}^1$. The magnitude of $\mathbf{DFT} : \mathbb{R}^1 \rightarrow \mathbb{C}^1$ of the sequence after applying this mapping is as follows:

![DFT Magnitude Plot]

The magnitude of the DFT at 10Hz is only 1.163575 (10\ log_{10}(1.163575) = 0.657944 \text{ dB}). There are exactly 2023 out of a total $\frac{N}{2} = 6000$ values that are greater than the DFT value at 10Hz (that value being 1.163575).\(^{46}\) Here again, the 10Hz component is effectively lost in the noise.

3. Suppose we next use the **QPSK DNA random variable** (Definition 3.5 page 83) to map the DNA sequence into the complex plane. The magnitude of $\mathbf{DFT} : \mathbb{C}^1 \rightarrow \mathbb{C}^1$ of the sequence after applying this mapping is as follows:

![DFT Magnitude Plot]

The DFT at 10Hz is 1.888671, or 10\log_{10}(1.888671) = 2.761563 \text{ dB}. There are exactly 343 out of a total $N = 6000$ values that are greater than the DFT value at 10Hz. (that value being 1.888671).\(^{47}\) Using this mapping it would be difficult to detect the subtle but significant 10Hz component.

4. Suppose we next use the $\mathbb{R}^4$ **DNA random variable** (Definition 3.12 page 83) to map the sequence into $\mathbb{R}^4$. The magnitude of $\mathbf{DFT} : \mathbb{R}^4 \rightarrow \mathbb{C}^1$ of the mapped sequence is as follows:

![DFT Magnitude Plot]

The magnitude of the DFT at 10Hz is 1.932042 (10\log_{10}(1.932042) = 2.860166 \text{ dB}). Besides itself and the DC component, there are only two out of a total $\frac{N}{2} = 6000$ samples that are greater or equal to this value.\(^{48}\) Thus, using the $\mathbb{R}^4$ mapping technique and subsequent analysis of this example, it is much simpler to detect the 10Hz oscillation.

**Example 3.22 (Fourier analysis of Ebola DNA sequence).**

1. Consider the Ebola DNA sequence described in Example 3.9 (page 81). DNA sequences commonly exhibit a strong DFT harmonic component at $\frac{2\pi}{6}$ radians.\(^{49}\)

2. Suppose we first use the **PAM DNA random variable** (Definition 3.10 page 83) to map the DNA sequence into $\mathbb{R}^1$. The magnitude of $\mathbf{DFT} : \mathbb{R}^1 \rightarrow \mathbb{C}^1$ of the sequence after applying this mapping is as follows:

\(^{46}\) See experiment log file “dnadft_12000m1200.xlg” generated by the program “ssp.exe”.

\(^{47}\) See experiment log file “dnadft_12000m1200.xlg” generated by the program “ssp.exe”.

\(^{48}\) See experiment log file “dnadft_12000m1200.xlg” generated by the program “ssp.exe”.

\(^{49}\) See experiment log file “dnadft_12000m1200.xlg” generated by the program “ssp.exe”.

The 4 largest values are at (0, 54.791926), (10, 1.932042), (4187, 1.962836), and (5147, 2.057553).\(^{50}\)

\(^{50}\) Galleani and Garello (2010) page 771
The component at $2^{\frac{j}{j}}$ is easy to pick out with a signal to noise ratio (SNR) of $10 \log_{10}(4.290296/1.123163) \approx 5.8$ dB. Here, the noise value 1.123163 is the RMS (root mean square) of the DFT magnitude sequence from $n = 1$ to $n = N/2 - 1$ computed as follows:

$$\sqrt{\frac{1}{N/2 - 1} \sum_{n=1}^{N/2 - 1} x_n^2}.$$

3. Suppose we next use the QPSK DNA random variable (Definition 3.5 page 83) to map the dna sequence into the complex plane. The magnitude of DFT : $\mathbb{C}^1 \to \mathbb{C}^1$ of the sequence after applying this mapping is as follows:

The component at $2^{\frac{j}{j}}$ is again easy to pick out with a signal to noise ratio (SNR) of $10 \log_{10}(6.412578/0.998659) \approx 8.1$ dB. Here, the noise value 0.998659 is the RMS of the DFT magnitude sequence from $n = 1$ to $n = N - 1$.

4. Suppose we next use the $\mathbb{R}^4$ DNA random variable (Definition 3.12 page 83) to map the sequence into $\mathbb{R}^4$. The magnitude of DFT : $\mathbb{R}^4 \to \mathbb{C}^4$ of the mapped sequence is as follows:

The component at $2^{\frac{j}{j}}$ is again easy to pick out with a signal to noise ratio (SNR) of $10 \log_{10}(3.944811/0.860665) \approx 6.6$ dB. Here, the noise value 0.860665 is the RMS of the DFT magnitude sequence from $n = 1$ to $n = N/2 - 1$.

5. In conclusion, for this application, there is only a small advantage to using the $\mathbb{R}^4$ mapping (item 4) versus the $\mathbb{R}^2$ mapping (item 2), and even a demonstrable disadvantage when compared to the $\mathbb{C}^1$ mapping (item 3).

3.3.4 Wavelet Analysis

In this section, we use what is in essence wavelet analysis, but yet is not truly wavelet analysis in the strict sense:
1. For starters, standard wavelets and their associated scaling functions are not sequences (Definition 1.41 page 23), but rather are functions with domain \( \mathbb{R} \) (not \( \mathbb{Z} \) or some convex subset of \( \mathbb{Z} \)).

2. While it is true that the celebrated Fast Wavelet Transform (FWT) does work internally with sequences (using filter banks),\(^{53}\) the FWT is actually defined to work on functions with domain \( \mathbb{R} \); and so the function to be analyzed by the FWT must first be sampled by a scaling function, which yields a sequence that can be processed by the filter banks.\(^{54}\)

3. Wavelet analysis is typically performed by translating the wavelet or scaling function by fixed amounts depending on the “scale” of the given wavelet. For example, a Haar wavelet of length 4000 would typically “jump” in offsets of 4000: 0, 4000, 8000, 12000, ….\(^{55}\) As one might imagine, this may be reason for concern if you are using this wavelet to perform edge detection (you might jump over and miss detecting the edge). In this section, wavelet sequences are translated by offsets of 1, making an edge harder to miss.

**Example 3.23** (statistical edge detection using Haar wavelet on non-stationary die sequence).

1. Suppose we have a length \( N = 12000 \) die sequence \((x_n)\) with the following distribution:

\[
P(x) = P(x_{-1}) = P(x_{-2}) = P(x_{-3}) = P(x_{-4}) = \frac{1}{6} \quad \text{for} \quad n \in [0 : 3999] \cup [8000 : 11999]
\]

\[
P(x) = P(x_{-1}) = P(x_{-2}) = P(x_{-3}) = P(x_{-4}) = \frac{1}{10} \quad \text{and} \quad P(x_{-3}) = \frac{1}{5} \quad \text{for} \quad n \in [4000 : 7999]
\]

That is, the distribution of the sequence is uniformly distributed in the first and last thirds, but biased towards \( \square \) in the middle third. In this example we use a simple statistical edge detector to try to find the statistical “edges” at 4000 and 8000. The edge detector here is a filter operation \( W \) (Definition 1.45 page 25) using a length 200 Haar wavelet sequence (Definition 1.51 page 25).\(^ {56}\)

2. Suppose we first use the PAM die random variable (Definition 3.4 page 82) to map the sequence of item (1) into \( \mathbb{R}^1 \). The magnitude of \( W : \mathbb{R}^1 \to \mathbb{C}^1 \) of the mapped sequence is as follows:\(^ {57}\)

![Image](image1.png)

We might expect to see strongest evidence of the edges at 4000 + 200/2 = 4100 and 8100. But looking at the above result, this is not apparent. In fact, there are a total of 10646 values that are greater than or equal to the value at location 4100 (that value being 0.015).\(^ {58}\)

3. Suppose we next use the QPSK die random variable (Definition 3.5 page 83) to map the die sequence into the complex plane. The magnitude of \( W : \mathbb{C}^1 \to \mathbb{C}^1 \) of the mapped sequence is as follows:

![Image](image2.png)

---


\(^{56}\) Empirical evidence due to Singh et al. (1997) suggests that the Haar wavelet performs better than several other common wavelets as an edge detector.

\(^{57}\) Note that the plot in item (2) has been down sampled by a factor of 10 for practical reasons of displaying the very large data set.

\(^{58}\) See experiment log file “diehaar_12000m4000_h200_1050.png” generated by the program “ssp.exe”.

2016 Jun 16 6:40 AM UTC A Thesis concerning symbolic sequence processing Daniel J. Greenhoe
Using this method, the edges are apparent. And the value of the peak at \( n = 4083 \) (with value 0.223383) is about \( 10 \log_{10}(0.223383/0.072741) \approx 4.9 \) dB above the noise floor. Here, the noise value 0.072741 is the RMS (see item (2) of Example 3.22 page 99) of the DFT magnitude sequence computed over the domain \( n = 200 \ldots N - 1 \).

4. Suppose we next use the \( \mathbb{R}^6 \) die random variable (Definition 3.7 page 83) to map the sequence into \( \mathbb{R}^6 \). The magnitude of \( W : \mathbb{R}^6 \rightarrow \mathbb{R}^6 \) of the mapped sequence is as follows:

![Graph showing magnitude of W vs frequency]

Using this method, the edges are also apparent. And the value of the peak at \( n = 4102 \) (with value 0.209165) is about \( 10 \log_{10}(0.209165/0.065768) \approx 5.0 \) dB above the noise floor. This is only a slight improvement over item (3).

**Example 3.24** (statistical edge detection using Haar wavelet on non-stationary artificial DNA sequence).

1. Suppose we have a length \( N \triangleq 12000 \) dna sequence \( (x_n) \) with the following distribution:

\[
P(\xi) = P(\varepsilon) = P(\zeta) = P(\varphi) = \frac{1}{4} \quad \text{for } n \in [0 : 3999] \cup [8000 : 11999]
\]

\[
P(\varepsilon) = P(\zeta) = P(\varphi) = \frac{27}{100} \quad \text{and } P(\xi) = \frac{49}{100} \quad \text{for } n \in [4000 : 7999]
\]

That is, the distribution of the sequence is uniformly distributed in the first and last thirds, but biased towards \( \varepsilon \) in the middle third. Just as in Example 3.23, we again use a filter operation \( W \) with length 200 Haar wavelet sequence as a simple statistical edge detector to try to locate the statistical “edges” at 4000 and 8000.

2. Suppose we first use the PAM DNA random variable (Definition 3.10 page 83) to map the DNA sequence into \( \mathbb{R}^1 \). The magnitude of \( W : \mathbb{R}^1 \rightarrow \mathbb{C}^1 \) of the mapped sequence is as follows:

![Graph showing magnitude of W vs frequency]

We might expect to see strongest evidence of the edges at or near \( 4000 + 200/2 = 4100 \) and 8100. In fact, the sequence does have peaks at 4087 (with value 0.185) and at 8087 (with value 0.230). The peak at 4087 is about \( 10 \log_{10}(0.185000/0.071355) \approx 4.1 \) dB above the noise floor. However, there are 103 other values not around the \( n = 4087 \) and \( n = 8087 \) peaks that are 0.185 or greater. These 102 values represent roughly 11 other peaks, each of which could trigger a “false positive” decision.

3. Suppose we next use the QPSK DNA random variable (Definition 3.5 page 83) to map the dna sequence into the complex plane. The magnitude of \( W : \mathbb{C}^1 \rightarrow \mathbb{C}^1 \) of the mapped sequence is

---

59 See experiment log file “diehaar_12000m4000_h200_1050.xlg” generated by the program “ssp.exe”.
60 Here the RMS noise value is computed over the domain \( n = 200 \ldots N - 1 \).
61 See experiment log file “diehaar_12000m4000_h200_1050.xlg” generated by the program “ssp.exe”.
62 Note that the plot in item (2) has been down sampled by a factor of 10 for practical reasons of displaying the very large data set.
63 See experiment log file “dnahaar_12000m4000_h200_1749.xlg” generated by the program “ssp.exe”.

---

**A thesis concerning symbolic sequence processing**

---

Daniel J. Greenhoe

2016 Jun 16 6:40AM UTC
as follows:

Using this method, the edges are apparent. And the value of the peak at \(n = 4086\) (with value 0.215870) is \(10 \log_{10}(0.215870/0.070068) \approx 4.9\) dB above the noise floor.\(^{63}\)

4. Suppose we next use the \(\mathbb{R}^4\) DNA random variable (Definition 3.12 page 83) to map the sequence into \(\mathbb{R}^4\). The magnitude of \(W : \mathbb{R}^4 \rightarrow \mathbb{C}^4\) of the mapped sequence is as follows:

Using this method, the edges are also apparent. And the value of the peak at \(n = 4096\) (with value 0.181246) is \(10 \log_{10}(0.181246/0.059594) \approx 4.8\) dB above the noise floor.\(^{64}\) Note that this is a slight decrease in performance as compared to item (3).

**Example 3.25** (Wavelet analysis of Phage Lambda DNA sequence).

1. Consider the Phage Lambda DNA sequence. It has a strong \(\mathbb{R}^1\) bias before \(n = 20000\) and a strong \(\mathbb{R}^4\) bias after,\(^{65}\) as demonstrated next by filtering the DNA sequence with a length 1600 Haar scaling sequence (Definition 1.50 page 25)—such a filtering operation acts as a kind of “sliding window” histogram of the DNA sequence.

2. Suppose we first use the PAM DNA random variable (Definition 3.10 page 83) to map the DNA sequence into \(\mathbb{R}^1\). The magnitude of the length 1600 Haar wavelet operation on the mapped sequence is as follows:

Note that it is very difficult to pick out the edge at 20000.

---

\(^{63}\) See experiment log file “dnahaar_12000m4000_h200_1749.xls” generated by the program “ssp.exe”.

\(^{64}\) See experiment log file “dnahaar_12000m4000_h200_1749.xls” generated by the program ”ssp.exe”.

\(^{65}\) Cristianini and Hahn (2007) page 14
3. Suppose we next use the QPSK DNA random variable (Definition 3.11 page 83) to map the DNA sequence into the complex plane. The length 1600 Haar wavelet operation on the mapped sequence is as follows:

If one did not know apriori that there was an edge at 20000, it would still be difficult to identify.

4. Suppose we next use the $\mathbb{R}^4$ DNA random variable (Definition 3.12 page 83) to map the DNA sequence into $\mathbb{R}^4$. Filtering the mapped sequence with a length 1600 Haar wavelet sequence results in the following:

Here there is a clear peak near 20000.

5. And here is the same analysis as used in item (4), but at scale 4000 (using a length 4000 Haar wavelet filter):

Again, the peak near 20000 is quite pronounced. However, at the low resolution scale (of 4000), it would be difficult to determine precisely where the statistical edge actually was.
APPENDIX A

LAGRANGE ARC DISTANCE

“The bottom line for mathematicians is that the architecture has to be right. In all the mathematics that I did, the essential point was to find the right architecture. It’s like building a bridge. Once the main lines of the structure are right, then the details miraculously fit. The problem is the overall design.”

Freeman Dyson (1923–), physicist and mathematician ¹

A.1 Introduction

A.1.1 The spherical metric

The spherical metric, or great circle metric, ² p, operates on the surface of a sphere with radius r centered at the origin (0, 0, ..., 0) in a linear space \( \mathbb{R}^N \), where “surface of a sphere...” is defined as all the points in \( \mathbb{R}^N \) that are a distance r from (0, 0, ..., 0) with respect to the Euclidean metric (Definition D.10 page 156). Thus, for any pair of points \((p, q)\) on the surface of this sphere, \((p, q)\) is in the domain of \( p \), and \( p, (p, q) \) is the “distance” between those points. However, if \( p \) and \( q \) are both in \( \mathbb{R}^N \) but are not on the surface of a common sphere centered at the origin, then \((p, q)\) is not in the domain of \( p \), and \( p, (p, q) \) is simply undefined.

In certain applications, however, it would be useful to have an extension \( d \) of the spherical metric \( p \) to the entire space \( \mathbb{R}^N \) (rather than just on a surface in \( \mathbb{R}^N \)). For example, for the points \( p \triangleq (0, 1) \) and \( q \triangleq (1, 0) \) (which are both on the surface of a common sphere in \( \mathbb{R}^2 \)), we would like \( d \) to be compatible with \( p \) such that \( d(p, q) = p(p, q) \). If \( r \triangleq (2, 0) \), then the pair \((p, r)\) is not in the domain of \( p \), but we still would like it to be in the domain of \( d \) such that \( d(p, r) \) is defined—and in this way \( d \) would be an extension of \( p \).

In this text, the Langrange arc distance is used in

¹ quote: Albers and Dyson (1994), page 20
the low pass filtering of a real die sequence
the low pass filtering of a spinner sequence
the high pass filtering of a weighted real die sequence
the high pass filtering of a weighted spinner sequence

A.1.2 Linear interpolation

This paper introduces an extension to the spherical metric based on a polar form of linear interpolation. Interpolation has a very long history with evidence suggesting that it extends possibly all the way back to the Babylonians living around 300BC.³

Linear interpolation between two points $p \triangleq (x_1, y_1)$ and $q \triangleq (x_2, y_2)$ in $\mathbb{R}^2$ is conveniently and intuitively expressed in a Cartesian coordinate system using what is commonly known as Lagrange interpolation (Definition E.1 page 173) in the form

$$y = y_1 \left( \frac{x - x_2}{x_1 - x_2} \right) + y_2 \left( \frac{x - x_1}{x_2 - x_1} \right).$$

Newton interpolation (Definition E.2 page 174) yields the same expression, but generally requires more “effort” (back substitution or matrix algebra):

$$y \triangleq \sum_{k=1}^{2} \sum_{m=1}^{k} \alpha_k \sum_{m=1}^{k} (x - x_m) = \alpha_1 [x - x_1] + \alpha_2 [(x - x_1) + (x - x_2)] = (\alpha_1 + \alpha_2)(x - x_1) + \alpha_2(x - x_2)$$

$$\text{Newton polynomial (Definition E.2)}$$

$$y_1 = \alpha_1 [x_1 - x_1] + \alpha_2 [(x_1 - x_1) + (x_1 - x_2)] \quad \Longrightarrow \quad \alpha_2 = \frac{y_1}{x_1 - x_2}$$

$$y_2 = \alpha_1 [x_2 - x_1] + \alpha_2 [(x_2 - x_1) + (x_2 - x_2)] \quad \Longrightarrow \quad \alpha_1 = \frac{y_1 + y_2}{x_2 - x_1}$$

$$y = \left[ y_1 \frac{y_1 + y_2}{x_2 - x_1} + \frac{y_1}{x_1 - x_2} \right] (x - x_1) + \left[ \frac{y_1}{x_1 - x_2} (x - x_2) = y_1 \left( \frac{x - x_2}{x_1 - x_2} \right) + y_2 \left( \frac{x - x_1}{x_2 - x_1} \right) \right.$$  

Of course the 2-point Lagrange interpolation/Newton interpolation polynomial can also be written in the familiar slope-intercept $y = mx + b$ form as

$$y = y_1 \left( \frac{x - x_2}{x_1 - x_2} \right) + y_2 \left( \frac{x - x_1}{x_2 - x_1} \right) = y_1 \left( \frac{x - x_1}{x_2 - x_1} \right) - y_1 \left( \frac{x - x_2}{x_2 - x_1} \right) = \frac{y_2(x - x_1) - y_1(x - x_2)}{x_2 - x_1}$$

$$= \left( \frac{y_2 - y_1}{x_2 - x_1} \right) x + \left( \frac{x_2 y_1 - x_1 y_2}{x_2 - x_1} \right)$$

A.1.3 Polar linear interpolation

Linear interpolation is often illustrated in terms of cartesian coordinates $(x, y)$. But there is no reason why the same principles cannot be used in terms of polar coordinates $(r(\theta), \theta)$. However care does need to be taken where $\theta$ may be interpreted to “jump” from $2\pi$ to 0 or from $-\pi$ to $\pi$.

³ Meijering (2002) page 320
Here is an expression for 2-point Lagrange interpolation/Newton interpolation in polar form:

\[ r(\theta) \triangleq r_p \left[ \frac{\theta - \theta_q}{\theta_p - \theta_q} \right] + r_q \left[ \frac{\theta - \theta_p}{\theta_q - \theta_p} \right] \quad \forall \theta \in [\theta_p : \theta_q] \]

Note the following:

1. The orientation of the axes in plane \( P \) is arbitrary, and that without loss of generality we can orient the axes such that \( p \) or \( q \) is on the positive \( x \)-axis and that the other point has a non-negative \( y \) value.

2. This means that the length of the arc between \( p \) at \( (r_p, \theta_p) \) and \( q \) at \( (r_q, \theta_q) \) under the original orientation is equal to the length of the arc between the points \( (r_p, 0) \) and \( (r_q, |\theta_p - \theta_q|) \) in the new orientation.

3. One important reason for the geometrical acrobatics here is that we don’t want to have to calculate the values for \( \theta_p \) and \( \theta_q \) in a plane \( P \) (which we don’t even immediately have an algebraic expression for anyways). But calculating the value \( \phi \triangleq |\theta_p - \theta_q| \) is quite straightforward because the “dot product” \( \langle p \mid q \rangle \) of \( p \) and \( q \) (which is very easy to calculate) in \( \mathbb{R}^N \) equals \( r_p r_q \cos \phi \) (and so \( \phi = \arccos \left( \frac{1}{r_p r_q} \langle p \mid q \rangle \right) \)).

4. Actually, \( \phi = |\theta_q - \theta_p| \), as demonstrated below:

\[
\phi \triangleq \arccos \left( \frac{1}{r_p r_q} \sum_{n=1}^{N} x_n y_n \right) \quad \text{by definition of } \phi \\
\triangleq \arccos \left( \frac{1}{r_p r_q} \langle p \mid q \rangle \right) \quad \text{a standard definition from the field of “linear algebra”} \\
= \arccos \left( \frac{1}{r_p r_q} [r_p r_q \cos |\theta_q - \theta_p|] \right) \quad \text{a standard result from the field of “linear algebra”} \\
= \left| \theta_q - \theta_p \right|, \quad 2\pi - \left| \theta_q - \theta_p \right| \quad \text{by definition of } \arccos(x) \text{ and } \cos(x) \\
= \left| \theta_q - \theta_p \right| \quad \text{by item (1)}
\]

5. Setting \( \theta_p = 0 \) and \( \theta_q = \phi \) yields the following:

\[
r(\theta) = r_p \left[ \frac{\theta - \theta_q}{\theta_p - \theta_q} \right] + r_q \left[ \frac{\theta - \theta_p}{\theta_q - \theta_p} \right] \quad \text{Langrange form (Definition E.1 page 173)}
\]

\[
= r_p \left[ \frac{\theta - \phi}{0 - \phi} \right] + r_q \left[ \frac{\theta - 0}{\phi - 0} \right] = \frac{-r_p \theta + r_q \phi + r_q \theta}{\phi} \\
= \left( \frac{r_q - r_p}{\phi} \right) \theta + r_p \quad \text{polar slope-intercept form}
\]
A.1.4 Distance in terms of polar linear interpolation arcs

This paper introduces a new function herein called, for better or for worse,⁴ the Lagrange arc distance (Definition A.1 page 108) \( d(p, q) \). It’s domain is the entire space \( \mathbb{R}^N \). It is an extension of the spherical metric, which only has as domain the surface of a sphere in \( \mathbb{R}^N \).

When \( p \) or \( q \) is at the origin, or when the polar angle \( \phi \) between \( p \) and \( q \) is 0, then the Lagrange arc distance \( d(p, q) \) is simply a \( \frac{1}{2} \) scaled Euclidean metric (Definition D.10 page 156). In all other cases, \( d(p, q) \) is the \( \frac{1}{2} \) scaled length of the Lagrange interpolation arc extending from \( p \) to \( q \).

An equation for the length of an arc in polar coordinates is⁵

\[
R(p, q) = \int_{\theta_p}^{\theta_q} \sqrt{r^2(\theta) + \left( \frac{dr}{d\theta} \right)^2} \, d\theta
\]

This integral may look intimidating. Later however, Theorem A.1 (page 111) demonstrates that it has an “easily” computable and straightforward solution only involving arithmetic operators (+, −, …), the absolute value function \( |x| \), the square root function \( \sqrt{x} \), and the natural log function \( \ln(x) \).

Finally, note that the extension does come at a cost—the Lagrange arc distance is not a metric (Definition D.7 page 153), but rather only a distance (Definition B.1 page 125, Theorem A.4 page 118). For more details about the impact of this cost, see Theorem B.1 (page 125).

A.2 Definition

Definition A.1. Let \( p \triangleq (x_1, x_2, \ldots, x_N) \) and \( q \triangleq (y_1, y_2, \ldots, y_N) \) be two points in the space \( \mathbb{R}^N \) with origin \((0, 0, \ldots, 0)\). Let

\[
\begin{align*}
  r_p &\triangleq \left( \sum_{n=1}^{N} x_n^2 \right)^{\frac{1}{2}} \\
  r_q &\triangleq \left( \sum_{n=1}^{N} y_n^2 \right)^{\frac{1}{2}} \\
  \phi &\triangleq \arccos \left( \frac{1}{r_p r_q} \sum_{n=1}^{N} x_n y_n \right)
\end{align*}
\]

(length of the arc \( \phi \) between \( p \) and \( q \))

The Lagrange interpolation polynomial

\[
r(\theta) \triangleq \left( \frac{r_q - r_p}{\phi} \right) \theta + r_p
\]

(magnitude of \( p \))

(magnitude of \( q \))

(angle between \( p \) and \( q \))

The Lagrange arc distance \( d(p, q) \) is defined as

\[
d(p, q) = \left\{ \begin{array}{ll}
\frac{r_p}{|r_p - r_q|} & \text{if } p = (0, 0, \ldots, 0) \text{ or } q = (0, 0, \ldots, 0) \text{ or } \phi = 0 \\
\frac{r_q}{|r_p - r_q|} R(p, q) & \text{otherwise}
\end{array} \right\}
\]

⁴“for better or for worse”: As already pointed out, Newton interpolation or simply the slope-intercept form \( y = mx + b \) of the line equation can with a little bit of effort give you the same equation as the 2-point Lagrange interpolation. So why not name the function \( d(p, q) \) of Definition A.1 “Newton arc distance”? Actually Newton published his interpolation method (for example in his 1711 "Methodus differentialis" [Newton (1711)]) long before Lagrange (Lagrange (1777)). But besides that, Lagrange was not really the first to discover what is commonly called “Lagrange interpolation”. The same result was actually published about 98 years earlier by Edward Waring (Waring (1779)). But in the end, the choice to use the name “Lagrange arc distance” has some justification in that it’s form arguably comes more readily using Lagrange interpolation than it does from Newton interpolation (which requires back substitution); and even though “Lagrange interpolation” probably should be called “Waring interpolation”, the fact is that it’s normally called “Lagrange interpolation”. So there is some motivation for the choice of the name. And “for better or for worse”, the function \( d(p, q) \) is herein called the “Lagrange arc distance”. …One last note: for a much fuller historical background of interpolation, see Meijering (2002).

⁵Stewart (2012) page 533 (Section 9.4 Areas and lengths in polar coordinates)
A.3 Calculation

The integral in Definition A.1 may look intimidating. However, Theorem A.1 (page 111) demonstrates that it has an “easily” computable and straightforward solution only involving arithmetic operators (+, −, …,), the absolute value function |x|, the square root function \( \sqrt{x} \), and the natural log function \( \ln(x) \). But first, a lemma (next) to help with the proof of Theorem A.1.

**Lemma A.1.** Let \( \sqrt{x} \in \mathbb{R}^a \) be the SQUARE ROOT function, and \( \ln(x) \triangleq \log_\varphi(x) \in \mathbb{R}^a \) be the NATURAL LOG function. Let \( \varepsilon \) be any given value in \( \mathbb{R}^a \).

\[
\begin{align*}
\{ 2ax + b + 2\sqrt{a(ax^2 + bx + c)} > 0 \} & \quad \implies \\
\int \frac{2ax + b}{4a} \sqrt{ax^2 + bx + c} \, dx &= \frac{2ax + b}{4a} \sqrt{ax^2 + bx + c} + \frac{4ac - b^2}{8a^{3/2}} \ln \left( 2ax + b + 2\sqrt{a(ax^2 + bx + c)} \right) + \varepsilon
\end{align*}
\]

**Proof:**

1. lemma: (first equality by the product rule, and the second equality by the chain rule)

\[
\frac{d}{dx} \left( \frac{2ax + b}{4a} \right) \sqrt{ax^2 + bx + c} = \left( \frac{2ax + b}{4a} \right) \left( \frac{d}{dx} \sqrt{ax^2 + bx + c} \right) + \left( \frac{d}{dx} \frac{2ax + b}{4a} \right) \left( \sqrt{ax^2 + bx + c} \right)
\]

\[
= \left( \frac{2ax + b}{4a} \right) \left( \frac{2ax + b}{2\sqrt{ax^2 + bx + c}} \right) + \left( \frac{2a}{4a} \right) \left( \sqrt{ax^2 + bx + c} \right)
\]

\[
= \frac{(2ax + b)^2 + 4a(ax^2 + bx + c)}{8a\sqrt{ax^2 + bx + c}}
\]

\[
= \frac{4a^2x^2 + 4axb + b^2 + 4a^2x^2 + 4abx + 4ac}{8a\sqrt{ax^2 + bx + c}}
\]

\[
= \frac{8a^2x^2 + 8abx + b^2 + 4ac}{8a\sqrt{ax^2 + bx + c}}
\]

2. lemma: If \( 2ax + b + 2\sqrt{a(ax^2 + bx + c)} > 0 \) then

\[
\frac{d}{dx} \left( \frac{4ac - b^2}{8a^{3/2}} \ln \left( 2ax + b + 2\sqrt{a(ax^2 + bx + c)} \right) \right)
\]

\[
= \left( \frac{4ac - b^2}{8a^{3/2}} \right) \left( \frac{d}{dx} \ln \left( 2ax + b + 2\sqrt{a(ax^2 + bx + c)} \right) \right)
\]

\[
= \left( \frac{4ac - b^2}{8a^{3/2}} \right) \left( \frac{2ax + b + 2\sqrt{a(ax^2 + bx + c)}}{2\sqrt{ax^2 + bx + c}} \right) \left( \frac{2a + 2\sqrt{a(2ax + b)}}{2\sqrt{ax^2 + bx + c}} \right) \quad \text{by linearity of } \frac{d}{dx}
\]

\[
= \left( \frac{4ac - b^2}{8a^{3/2}} \right) \left( \frac{2a\sqrt{ax^2 + bx + c} + \sqrt{a(2ax + b)}}{2a\sqrt{ax^2 + bx + c} + \sqrt{a(2ax + b)}} \right) \quad \text{by chain rule}
\]

\[
= \left( \frac{4ac - b^2}{8a^{3/2}} \right) \left( \frac{2a\sqrt{ax^2 + bx + c} + \sqrt{a(2ax + b)}}{2a\sqrt{ax^2 + bx + c} + \sqrt{a(2ax + b)}} \right)
\]

\[
= \frac{8a\sqrt{ax^2 + bx + c} + 2a\sqrt{ax^2 + bx + c}}{8a\sqrt{ax^2 + bx + c} + 2a\sqrt{ax^2 + bx + c}}
\]

\[
= \frac{4ac - b^2}{8a\sqrt{ax^2 + bx + c}}
\]

---

6 Gradshteyn and Ryzhik (2007) page 94 (2.25 Forms containing \( \sqrt{a + bx + cx^2} \), 2.26 Forms containing \( \sqrt{a + bx + cx^2} \) and integral powers of \( x \); Jeffrey (1995) page 160 (4.3.4 Integrands containing \( (a + bx + cx^2)^5 \), Jeffrey and Dai (2008) pages 172–173 (4.3.4 Integrands containing \( (a + bx + cx^2)^5 \)).
3. Lemma: If \( (2ax + b + 2\sqrt{a(ax^2 + bx + c)}) < 0 \) then

\[
\frac{d}{dx} \left[ \frac{4ac - b^2}{8a^{3/2}} \ln \left(2ax + b + 2\sqrt{a(ax^2 + bx + c)}\right) \right] = \left( \frac{4ac - b^2}{8a^{3/2}} \right) \left( \frac{1}{-2ax - b - 2\sqrt{a(ax^2 + bx + c)}} \right) \left( 2a + \frac{2\sqrt{a(2ax + b)}}{2\sqrt{ax^2 + bx + c}} \right) \]

by linearity of \( \frac{d}{dx} \)

\[
= \frac{(4ac - b^2)}{8a^{3/2}} \left[ \frac{2a}{\sqrt{ax^2 + bx + c}} + c + \sqrt{a(2ax + b)} \right]
\]

by chain rule

\[
= \frac{8a^{3/2}}{-2ax - b - 2\sqrt{a(ax^2 + bx + c)}} \sqrt{ax^2 + bx + c}
\]

\[
= -\frac{(4ac - b^2)}{8a\sqrt{ax^2 + bx + c}} \left[ \sqrt{a(2ax + b)} + 2a\sqrt{ax^2 + bx + c} \right]
\]

\[
= \frac{8a\sqrt{ax^2 + bx + c}}{8a\sqrt{ax^2 + bx + c}} \left[ \sqrt{a(2ax + b)} + 2a\sqrt{ax^2 + bx + c} \right]
\]

4. Complete the proof: If \( (2ax + b + 2\sqrt{a(ax^2 + bx + c)}) > 0 \) then

\[
\frac{d}{dx} \left[ \frac{2ax + b}{4a} \sqrt{ax^2 + bx + c} + \frac{4ac - b^2}{8a^{3/2}} \ln \left(2ax + b + 2\sqrt{a(ax^2 + bx + c)}\right) \right] = \frac{d}{dx} \left[ \left( \frac{2ax + b}{4a} \right) \sqrt{ax^2 + bx + c} \right] + \frac{d}{dx} \left[ \left( \frac{4ac - b^2}{8a^{3/2}} \right) \ln \left(2ax + b + 2\sqrt{a(ax^2 + bx + c)}\right) \right]
\]

by item (1)

\[
= \frac{8a^2x^2 + 8abx + b^2 + 4ac}{8a\sqrt{ax^2 + bx + c}} + \frac{4ac - b^2}{8a\sqrt{ax^2 + bx + c}} + \frac{4ac - b^2}{8a\sqrt{ax^2 + bx + c}} \ln \left(2ax + b + 2\sqrt{a(ax^2 + bx + c)}\right)
\]

by item (2)

\[
= \frac{8a(ax^2 + bx + c)}{8a\sqrt{ax^2 + bx + c}} = \sqrt{ax^2 + bx + c}
\]

5. Note that simply forcing the agreement of \( \ln \) to be positive as in\(^7\)

\[
\frac{2ax+b}{4a} \sqrt{ax^2 + bx + c} + \frac{4ac-b^2}{8a^{3/2}} \ln \left(2ax + b + 2\sqrt{a(ax^2 + bx + c)}\right) + \epsilon
\]

is not a solution to \( \int \sqrt{ax^2 + bx + c} \ dx \) when \( (2ax + b + 2\sqrt{a(ax^2 + bx + c)}) < 0 \):

\[
\frac{d}{dx} \left[ \frac{2ax + b}{4a} \sqrt{ax^2 + bx + c} + \frac{4ac - b^2}{8a^{3/2}} \ln \left(2ax + b + 2\sqrt{a(ax^2 + bx + c)}\right) \right] = \frac{d}{dx} \left[ \left( \frac{2ax + b}{4a} \right) \sqrt{ax^2 + bx + c} \right] + \frac{d}{dx} \left[ \left( \frac{4ac - b^2}{8a^{3/2}} \right) \ln \left(2ax + b + 2\sqrt{a(ax^2 + bx + c)}\right) \right]
\]

by item (1)

\[
= \frac{8a^2x^2 + 8abx + b^2 + 4ac}{8a\sqrt{ax^2 + bx + c}} + \frac{4ac - b^2}{8a\sqrt{ax^2 + bx + c}} \ln \left(2ax + b + 2\sqrt{a(ax^2 + bx + c)}\right)
\]

by item (3)

\[
= \frac{8a(ax^2 + bx + c)}{8a\sqrt{ax^2 + bx + c}} = \frac{8a(ax^2 + bx + c) + 2b^2 - 8ac}{8a\sqrt{ax^2 + bx + c}}
\]

\(^7\)The solution \( \ln \left[ \cdots \right] \) is used in \(^7\) Jeffrey (1995) page 160 and \(^7\) Jeffrey and Dai (2008) pages 172–173.
\[ = \sqrt{ax^2 + bx + c} + \frac{2b^2 - 8ac}{8a\sqrt{ax^2 + bx + c}} \neq \sqrt{ax^2 + bx + c} \quad \text{for } b^2 \neq 4ac \]

6. Note further that constraining \( a > 0 \) is also not a solution\(^8\) because it does not guarantee that the argument \( u \) of \( \ln(u) \) will be positive. Take for example \( a = 1, b = -3, c = 3 \) and \( x = 1 \). Then
\[
2ax + b + 2\sqrt{a(x^2 + bx + c)} = 2\cdot1\cdot1 - 3 + 2\sqrt{1(1 - 3 \cdot 1 + 3)} = -1 < 0.
\]

\[\square\]

**Theorem A.1.** Let \( R(p, q), r_p, r_q, \) and \( \phi \) be as defined in Definition A.1 (page 108). Let \( \rho \triangleq r_q - r_p \). If \( r_p \neq 0, r_q \neq 0 \) and \( \phi \neq 0 \) then
\[
R(p, q) = \frac{r_q \sqrt{(r_q \phi)^2 + \rho^2} - r_p \sqrt{(r_p \phi)^2 + \rho^2}}{2\rho} + \frac{|\rho|}{2\phi} \left[ \ln \left( |\rho| \sqrt{(r_q \phi)^2 + \rho^2} \right) - \ln \left( |\rho| \sqrt{(r_p \phi)^2 + \rho^2} \right) \right]
\]

\[\square\] Proof:

1. Let \( \gamma \triangleq r_p \phi \).

2. Lemmas:

\[
\rho \phi + \gamma = (r_q - r_p) \phi + r_p \phi = r_q \phi
\]
\[
\rho^2 \phi^2 + 2\gamma \phi + (\gamma^2 + \rho^2) = (r_q - r_p)^2 \phi^2 + 2(r_q - r_p)(r_q \phi) \phi + (r_q \phi)^2 + \rho^2
\]
\[
= (r_q^2 + r_p^2 - 2r_q r_p \phi) \phi^2 + 2(r_q - r_p)(r_q \phi) \phi + (r_q \phi)^2 + \rho^2
\]
\[
= (r_q^2 + r_p^2 \phi^2 - 2(r_q \phi)^2 + (r_q \phi)^2 + \rho^2
\]
\[
= (r_q \phi)^2 + \rho^2
\]

3. Lemma: \( 2\rho^2 \theta + 2\rho \gamma + 2\sqrt{\rho^2(\rho^2 \theta^2 + 2\rho \gamma \theta + (\gamma^2 + \rho^2))} > 0 \). Proof:

\[
2\rho^2 \theta + 2\rho \gamma + 2\sqrt{\rho^2(\rho^2 \theta^2 + 2\rho \gamma \theta + (\gamma^2 + \rho^2))}
\]
\[
> 2\rho^2 \theta + 2\rho \gamma + 2\sqrt{\rho^2 \gamma^2}
\]
\[
= 2\rho^2 \theta + 2\rho \gamma + 2|\rho|\gamma
\]
\[
= 2\rho^2 \theta + 2\gamma(\rho + |\rho|)
\]
\[
\geq 0
\]

4. Completing the proof...

\[
R(p, q) = \int_{\theta=\phi}^{\theta=0} \sqrt{r^2(\theta) + \left( \frac{dr}{d\theta} \right)^2} \, d\theta \quad \text{by def. of } R(p, q)
\]
\[
= \int_{0}^{\phi} \sqrt{\left( \frac{r_q - r_p}{\phi} \right)^2 + r_p^2} \, d\theta \quad \text{by def. of } r(\theta)
\]
\[
= \int_{0}^{\phi} \sqrt{\left( \frac{r_q - r_p}{\phi} \theta + r_p \phi \right)^2 + \left( r_q - r_p \right)^2} \, d\theta
\]
\[
= \frac{1}{\phi} \int_{0}^{\phi} \sqrt{(r_q - r_p)^2 \theta^2 + 2(r_q - r_p)(r_q \phi) \theta + (r_q \phi)^2} + (r_q - r_p)^2 \, d\theta
\]

---

\(^8\) The \( a > 0 \) constraint is used in Gradshteyn and Ryzhik (2007) page 94
\[
\frac{1}{\phi} \int_0^\phi \sqrt{\frac{\rho^2 \theta^2 + 2\rho \gamma (\gamma^2 + \rho^2)}{\sin \theta}} \, d\theta
\]

\[
= \frac{1}{\phi} \left[ \frac{2\rho^2 \theta + 2\rho \gamma}{4\rho^2} \sqrt{\rho^2 \theta^2 + 2\rho \gamma (\gamma^2 + \rho^2)} \right]_{\theta=0}^{\theta=\phi}
\]

(by item (3) and Lemma A.1)

\[
= \frac{1}{\phi} \left[ \frac{\rho \theta + \gamma}{2\rho} \sqrt{\rho^2 \theta^2 + 2\rho \gamma (\gamma^2 + \rho^2)} + \frac{\rho \theta}{2\rho} \ln \left( 2\rho \theta + 2\rho \gamma + 2|\rho| \sqrt{\rho^2 \theta^2 + 2\rho \gamma (\gamma^2 + \rho^2)} \right) \right]_{\theta=0}^{\theta=\phi}
\]

\[
= \frac{1}{\phi} \left[ \frac{(\rho \theta + \gamma) \sqrt{\rho^2 \theta^2 + 2\rho \gamma (\gamma^2 + \rho^2)}}{2\rho \phi} + \frac{|\rho|}{2\phi} \ln \left( 2\rho \phi + 2\rho \gamma + 2|\rho| \sqrt{\rho^2 \phi^2 + 2\rho \gamma (\gamma^2 + \rho^2)} \right) \right]
\]

\[
- \frac{\gamma \sqrt{\phi^2 + \rho^2} + \frac{|\rho|}{2\phi} \ln \left( 2\rho \phi + 2|\rho| \sqrt{\phi^2 + \rho^2} \right)}{2\rho \phi}
\]

\[
r_q \phi \sqrt{(r_q \phi)^2 + \rho^2} - r_p \phi \sqrt{(r_p \phi)^2 + \rho^2}
\]

\[
= \frac{2\rho \phi}{2\phi} \ln(2) + \ln \left( \rho^2 \phi + \rho \gamma + |\rho| \sqrt{(r_q \phi)^2 + \rho^2} \right) - ln(2) - \ln \left( \rho \gamma + |\rho| \sqrt{\phi^2 + \rho^2} \right)
\]

(by item (2))

\[
r_q \sqrt{(r_q \phi)^2 + \rho^2} - r_p \sqrt{(r_p \phi)^2 + \rho^2}
\]

\[
= \frac{2\rho}{2\phi} \left[ \ln \left( r_q \rho \phi + |\rho| \sqrt{(r_q \phi)^2 + \rho^2} \right) - \ln \left( r_p \rho \phi + |\rho| \sqrt{(r_p \phi)^2 + \rho^2} \right) \right]
\]

\[
= \frac{r_q - r_p}{2\rho}
\]

A.4 Properties

A.4.1 Arc function \( R(p,q) \) properties

If we really want the \textit{Langrange arc distance} \( d(p,q) \) to be an \textit{extension} of the \textit{spherical metric}, then \( R(p,q) \) must equal \( r_p \phi \) (the arc length between \( p \) and \( q \) on a circle centered at the origin) when \( r_p = r_q \). This is in fact the case, as demonstrated next.

\textbf{Proposition A.1 (R(p, q) on spherical surface).} Let \( R(p,q), r_p, r_q, \) and \( \phi \) be defined as in Definition A.1.

\[
\left\{ \begin{array}{l}
 r_p = r_q \\
 \phi \neq 0
\end{array} \right\} \quad \implies \quad \{ R(p,q) = r_p \phi \}
\]

\textit{Proof:}

1. lemma:

\[
\lim_{\rho \to 0} \frac{r_q \sqrt{(r_q \phi)^2 + \rho^2} - r_p \sqrt{(r_p \phi)^2 + \rho^2}}{2\rho} = \lim_{\rho \to 0} \frac{r_q \sqrt{(r_q \phi)^2 + 0} - r_p \sqrt{(r_p \phi)^2 + 0}}{2\rho} = \lim_{\rho \to 0} \frac{\phi}{2} \frac{r_q^2 - r_p^2}{r_q - r_p} = \lim_{\rho \to 0} \left( \frac{\phi}{2} \frac{r_q - r_p}{r_q + r_p} \right)
\]

\[
= \frac{r_p \phi}{2}
\]
2. lemma:

\[
\lim_{\rho \to 0} \frac{\rho}{2\phi} \left[ \ln \left( r_q \rho \phi + |\rho| \sqrt{(r_q \phi)^2 + \rho^2} \right) - \ln \left( r_p \rho \phi + |\rho| \sqrt{(r_p \phi)^2 + \rho^2} \right) \right]
\]

\[= \lim_{\rho \to 0} \frac{\rho}{2\phi} \left[ \ln \left( r_q \rho \phi + \rho \sqrt{(r_q \phi)^2} \right) - \ln \left( r_p \rho \phi + \rho \sqrt{(r_p \phi)^2} \right) \right]\]

\[= \lim_{\rho \to 0} \frac{\rho}{2\phi} \left[ \ln (\rho) + \ln \left( r_q \phi + \sqrt{(r_q \phi)^2} \right) - \ln (\rho) - \ln \left( r_p \phi + \sqrt{(r_p \phi)^2} \right) \right]\]

\[= \lim_{\rho \to 0} \frac{\rho}{2\phi} \left[ + \ln \left( r_q \phi + \sqrt{(r_q \phi)^2} \right) - \ln \left( r_p \phi + \sqrt{(r_p \phi)^2} \right) \right] = 0\]

3. lemma:

\[
\lim_{\rho \to 0} \frac{\rho}{2\phi} \left[ \ln \left( -r_q \rho \phi + |\rho| \sqrt{(r_q \phi)^2 + \rho^2} \right) - \ln \left( -r_p \rho \phi + |\rho| \sqrt{(r_p \phi)^2 + \rho^2} \right) \right]
\]

\[= \lim_{\rho \to 0} \frac{\rho}{2\phi} \left[ \ln \left( -r_q \rho \phi + \rho \sqrt{(r_q \phi)^2} \right) - \ln \left( -r_p \rho \phi + \rho \sqrt{(r_p \phi)^2} \right) \right]\]

\[= \lim_{\rho \to 0} \frac{\rho}{2\phi} \left[ \ln (\rho) + \ln \left( -r_q \phi + \sqrt{(r_q \phi)^2} \right) - \ln (\rho) - \ln \left( -r_p \phi + \sqrt{(r_p \phi)^2} \right) \right]\]

\[= \lim_{\rho \to 0} \frac{\rho}{2\phi} \left[ \ln \left( -r_q \phi + \sqrt{(r_q \phi)^2} \right) - \ln \left( -r_p \phi + \sqrt{(r_p \phi)^2} \right) \right] = 0\]

4. lemma: By item (2), item (3), and by continuity ...

\[
\lim_{\rho \to 0} \frac{\rho}{2\phi} \left[ \ln \left( r_q \rho \phi + |\rho| \sqrt{(r_q \phi)^2 + \rho^2} \right) - \ln \left( r_p \rho \phi + |\rho| \sqrt{(r_p \phi)^2 + \rho^2} \right) \right] = 0
\]

5. Completing the proof ...

\[
\lim_{\rho \to 0} R(\rho, q) = \lim_{\rho \to 0} \frac{r_q \sqrt{(r_q \phi)^2 + \rho^2} - r_p \sqrt{(r_p \phi)^2 + \rho^2}}{2\rho} + \frac{\rho}{2\phi} \ln \left( r_q \rho \phi + |\rho| \sqrt{(r_q \phi)^2 + \rho^2} \right) - \ln \left( r_p \rho \phi + |\rho| \sqrt{(r_p \phi)^2 + \rho^2} \right) \] by Theorem A.1

\[= 0 + \lim_{\rho \to 0} \frac{\rho}{2\phi} \left[ \ln \left( r_q \rho \phi + |\rho| \sqrt{(r_q \phi)^2 + \rho^2} \right) - \ln \left( r_p \rho \phi + |\rho| \sqrt{(r_p \phi)^2 + \rho^2} \right) \right] \] by item (1)

\[= 0 + 0 \] by item (4)

\[= 0 \]

\[\Rightarrow \]

Later in Theorem A.3 (page 117), we want to prove that the Langrange arc distance (Definition A.1 on page 108) d(p, q) is indeed, as its name suggests, a distance function (Definition B.1 on page 125). Proposition A.2 (symmetry) and Proposition A.4 (positivity) will help. Meanwhile, Proposition A.4 will itself receive help from Proposition A.3 (monotonicity).

**Proposition A.2** (symmetry of R). Let R(p, q) be defined as in Definition A.1 (page 108).

\[ R(p, q) = R(q, p) \quad \forall p, q \in \mathbb{R}^n \quad (\text{symmetric}) \]

**Proof:**

1. dummy variable: Let \( \mu \triangleq \phi - \theta \) which implies \( \theta = \phi - \mu \) and \( d\theta = -d\mu \).
2. lemma:
\[
r(\mu; q, p) = r(\phi - \theta; q, p) = \frac{r_p - r_q}{\phi} \phi \theta + r_q = (\frac{r_q - r_p}{\phi}) \phi \theta + (r_p - r_q) + r_q = (\frac{r_q - r_p}{\phi}) \phi \theta + r_p = r(\theta; p, q)
\]

3. Completing the proof...
\[
R(p, q) = \int_{\theta=0}^{\theta=\phi} \sqrt{r^2(\theta; p, q) + \left(\frac{dr(\theta; p, q)}{d\theta}\right)^2} \, d\theta
\]
\[
= \int_{\phi - \mu=0}^{\phi - \mu=\phi} \sqrt{r^2(\mu; q, p) + \left(\frac{dr(\mu; q, p)}{d\mu}\right)^2} \, d\mu \quad \text{by item (1) and item (2)}
\]
\[
= - \int_{\mu=0}^{\mu=\phi} \sqrt{r^2(\mu; q, p) + \left(\frac{dr(\mu; q, p)}{d\mu}\right)} \, d\mu
\]
\[
= \int_{\mu=0}^{\mu=\phi} \sqrt{r^2(\mu; q, p) + \left(\frac{dr(\mu; q, p)}{d\mu}\right)} \, d\mu \quad \text{by the Second Fundamental Theorem of Calculus}\]
\[
= R(q, p)
\]

Proposition A.3 (monotonicity of R). Let R(p, q) and \( \phi \) be defined as in Definition A.1 (page 108). Let \( \phi_1 \) be the polar angle between the point pair \( (p_1, q_1) \) in \( \mathbb{R}^N \) and Let \( \phi_2 \) the polar angle between the point pair \( (p_2, q_2) \) in \( \mathbb{R}^N \).
\[
\{ \phi_1 < \phi_2 \} \implies \{ R(p_1, q_1) < R(p_2, q_2) \} \quad \forall \phi_1, \phi_2 \in (0; \pi) \quad \text{(strictly monotonically increasing in } \phi)\]

\[\begin{aligned}
\text{PROOF:} \\
\frac{d}{d\phi} R(p, q) &= \frac{d}{d\phi} \int_{\theta=0}^{\phi} \sqrt{r^2(\theta) + \left(\frac{dr(\theta)}{d\theta}\right)^2} \, d\theta \quad \text{by Definition A.1 (page 108)} \\
&= \frac{d}{d\phi} \int_{\theta=0}^{\phi} \sqrt{\left(\frac{r_q - r_p}{\phi}\right)^2 + \left(\frac{r_q - r_p}{\phi}\right)^2} \, d\theta \\
&= \sqrt{\left(\frac{r_q - r_p}{\phi}\right)^2 + \left(\frac{r_q - r_p}{\phi}\right)^2} \quad \text{by the First Fundamental Theorem of Calculus}\] \\
&= \sqrt{r_q^2 + \left(\frac{r_q - r_p}{\phi}\right)^2} > 0 \\
\implies R(p, q) \text{ is strictly monotonically increasing in } \phi
\end{aligned}\]

Proposition A.4 (positivity of R). Let R(p, q) and \( \phi \) be defined as in Definition A.1 (page 108).
\[
\{ \phi \in (0; \pi) \} \implies \{ R(p, q) > 0 \} \quad \forall \mu, q \in \mathbb{R}^N \quad \text{(POSITIVE)}
\]

\[\begin{aligned}
\text{by Hijab (2016) page 170 (Theorem 4.4.3 Second Fundamental Theorem of Calculus),} \\
\text{by Amann and Escher (2008) page 31 (Theorem 4.13 The second fundamental theorem of calculus)}
\end{aligned}\]

\[\begin{aligned}
\text{by Schechter (1996) page 674 (25.15), by Haaser and Sullivan (1991) page 218}
\end{aligned}\]
\[ \frac{d}{d\phi} R(p, q) \triangleq \int_0^\phi \sqrt{r^2(\theta) + \left[ \frac{dr(\theta)}{d\theta} \right]^2} \, d\theta \]

by Definition A.1 (page 108)

\[ = \int \sqrt{r^2(\theta) + \left[ \frac{dr(\theta)}{d\theta} \right]^2} \, d\theta \bigg|_0^\phi - \int \sqrt{r^2(\theta) + \left[ \frac{dr(\theta)}{d\theta} \right]^2} \, d\theta \bigg|_0^\phi \]

by the Second Fundamental Theorem of Calculus

\[ > 0 \]

by Proposition A.3 (page 114)

For the sake of continuity at the origin of \( \mathbb{R}^N \), one might hope that it doesn’t matter which “direction” the points \( p \) or \( q \) approach the origin when computing the limit of \( R(p, q) \). This however is not the case, as demonstrated next and illustrated to the right and in Example A.1 (page 116). In fact, the limits very much depend on \( \phi \)...resulting in a discontinuity at the origin, as demonstrated in Theorem A.2 (page 117).

**Proposition A.5** (limit cases of \( R \)). Let \( R(p, q), r_p, r_q, \) and \( \phi \) be defined as in Definition A.1 (page 108).

\[
\lim_{r_q \to 0} R(p, q) = \frac{r_p}{2} \left[ \phi^2 + 1 + \ln \left( \phi + \sqrt{\phi^2 + 1} \right) \right] \quad \forall p, q \in (0, 0, \ldots, 0, \phi \neq 0) \quad (p \text{ approaching origin})
\]

\[
\lim_{r_q \to 0} R(p, q) = \frac{r_p}{2} \left[ \phi^2 + 1 + \ln \left( \phi + \sqrt{\phi^2 + 1} \right) \right] \quad \forall p, q \in (0, 0, \ldots, 0, \phi \neq 0) \quad (q \text{ approaching origin})
\]

**Proof:**

\[
\lim_{r_p \to 0} R(p, q)
\]

\[
= \lim_{r_p \to 0} \frac{r_q \sqrt{(r_q \phi)^2 + r_p^2} - r_p \sqrt{(r_p \phi)^2 + r_q^2}}{2 \rho} + \left| \frac{\rho}{2\phi} \ln \left( r_q \rho \phi + |\rho| \sqrt{(r_q \phi)^2 + r_p^2} \right) - \ln \left( r_p \rho \phi + |\rho| \sqrt{(r_p \phi)^2 + r_q^2} \right) \right|
\]

by Theorem A.1 (page 111)

\[
= \frac{r_q \sqrt{(r_q \phi)^2 + r_q^2} - 0}{2 \rho} + \frac{r_q}{2\phi} \ln \left( r_q \sqrt{(r_q \phi)^2 + r_q^2} + \phi \sqrt{\phi^2 + 1} \right) - \ln \left( 0 + |r_q| \sqrt{0 + r_q^2} \right) \quad \text{by limit operation}
\]

\[
= \frac{r_q}{2} \left[ \phi^2 + 1 + \ln \left( \phi + \sqrt{\phi^2 + 1} \right) - \ln \left( \phi + \sqrt{\phi^2 + 1} \right) \right]
\]

\[
= \frac{r_q}{2} \left[ \phi^2 + 1 + \frac{\ln \left( \phi + \sqrt{\phi^2 + 1} \right)}{\phi} \right]
\]

\[
= \lim_{r_q \to 0} R(p, q) = \lim_{r_q \to 0} R(q, p) \quad \text{by Proposition A.2}
\]

\[
= \frac{r_p}{2} \left[ \phi^2 + 1 + \frac{\ln \left( \phi + \sqrt{\phi^2 + 1} \right)}{\phi} \right] \quad \text{by previous result}
\]
Example A.1. Let $R(p, q), \phi$, and $r_q$ be defined as in Definition A.1.

- If $\phi = 0$, then \[
\lim_{p \to 0} R(p, (r_q, 0)) = r_q
\]
- If $\phi = \pi/2$, then \[
\lim_{p \to 0} R(p, (r_q, 0)) = r_q \times (1.323652 \ldots) \approx 1.3 r_q
\]
- If $\phi = \pi$, then \[
\lim_{p \to 0} R(p, (r_q, 0)) = r_q \times (1.944847 \ldots) \approx 1.9 r_q
\]

\[\Box\]

A.4.2 Distance function $d(p,q)$ properties

The *Langrange arc distance* $d(p,q)$ is defined in two parts: one part being the Euclidean distance $\sqrt{r_q^2 - r_p^2}$ and the second part the length of the arc $\frac{\phi}{r} R(p,q)$. There is risk in creating a multipart definition...with the possible consequences being discontinuity at the boundary of the parts. Proposition A.6 (next) demonstrates that when $r_p \neq 0$ and $r_q \neq 0$, there is continuity as $\phi \to 0$. However, Theorem A.3 (page 117) demonstrates that in general for values of $\phi > 0$, $d(p, q)$ is discontinuous at the origin.

Proposition A.6. Let $R(p, q), r_p, r_q, \phi$, and $(0,0,\ldots,0)$ be defined as in Definition A.1 (page 108).

(A) \[\lim_{\phi \to 0} R(p, q) = \left| r_q - r_p \right| = \pi d(p, q) \quad \text{when } \phi = 0\]

(B) \[\lim_{\phi \to 0} R(p, q) = \lim_{r_q \to 0} \lim_{r_p \to 0} R(p, q) = r_q = \pi d((0,0,\ldots,0), q)\]

(C) \[\lim_{\phi \to 0} R(p, q) = \lim_{r_q \to 0} \lim_{r_p \to 0} R(p, q) = r_p = \pi d(p, (0,0,\ldots,0))\]

\[\Box\]
\[ r_q - r_p \] = \pi \, d(p, q)_{|\phi=0} \] 

by Definition A.1 (page 108)

\[
\lim_{\phi \to 0} \lim_{r_q \to 0} R(p, q) = \lim_{\phi \to 0} \left[ \frac{r_q}{2} \left( \sqrt{\phi^2 + 1} + \ln \left( \phi + \sqrt{\phi^2 + 1} \right) \right) \right] 
\]

by Proposition A.5 (page 115)

\[
= \frac{r_q}{2} \left[ 1 + \lim_{\phi \to 0} \frac{\ln \left( \phi + \sqrt{\phi^2 + 1} \right)}{\phi} \right] = \frac{r_q}{2} \left[ 1 + 1/1 \right] = r_q = \pi d((0, 0, \ldots, 0), q) 
\]

by \text{L'Hôpital's rule}

by Proposition A.2 (page 113)

\[
= r_p = \pi d(p, (0, 0, \ldots, 0)) \] 

by previous result

\[
= r_q = \pi d((0, 0, \ldots, 0), q) \] 

by (A)

\[
= r_p = \pi d(p, (0, 0, \ldots, 0)) \] 

by (A)

\(\therefore\)

**Theorem A.2.** Let the LAGRANGE ARC DISTANCE \(d(p, q)\) and origin be defined as in Definition A.1. The function \(d(p, q)\) is DISCONTINUOUS at the origin of \(\mathbb{R}^N\), but is CONTINUOUS everywhere else in \(\mathbb{R}^N\).

**Proof:**

1. Proof for when \(p\) and \(q\) are not at the origin and \(\phi \neq 0\):
   
   (a) In this case, \(d(p, q) = 1/\phi R(p, q)\).
   
   (b) \(R(p, q)\) is continuous everywhere in its domain because its solution, as given by Theorem A.1 (page 111), consists entirely of continuous functions such as \(\ln(x), |x|, \) etc.
   
   (c) Therefore, in this case, \(d(p, q)\) is also continuous.

2. Proof for when \(p\) and \(q\) are not at the origin and \(\phi = 0\):
   
   This follows from (A) of Proposition A.6 (page 116).

3. Proof for discontinuity at origin: This follows from Proposition A.5 (page 115), where it is demonstrated that the limit of \(R(p, q)\) is very much dependent on the “direction” from which \(p\) or \(q\) approaches the origin. For an illustration of this concept, see Example A.1 (page 116).

\(\therefore\)

**Theorem A.3.** Let \(d(p, q)\) be LAGRANGE ARC DISTANCE (Definition A.1 page 108). The function \(d(p, q)\) is a DISTANCE FUNCTION (Definition B.1 page 125). In particular,

1. \(d(p, q) \geq 0\) \(\forall p, q \in \mathbb{R}^N\) (NON-NEGATIVE) and
2. \(d(p, q) = 0 \iff p = q\) \(\forall p, q \in \mathbb{R}^N\) (NONDEGENERATE) and
3. \(d(p, q) = d(q, p)\) \(\forall p, q \in \mathbb{R}^N\) (SYMMETRIC)
The Langrange arc distance is not a metric because in general the triangle inequality property does not hold (next theorem). Furthermore, the Langrange arc distance does not induce a norm because it is not translation invariant (the translation invariant property is a necessary condition for a metric to induce a norm, Theorem D.5 page 152), and balls in a Langrange arc distance space are in general not convex (balls are always convex in a normed linear space, Theorem D.4 page 150). For more details about distance spaces, see Appendix B (page 123).

Theorem A.4. In the Lagrange arc distance space \((X, d)\) over a field \(\mathbb{F}\)

1. \(d(p, r) \neq d(p, q) + d(q, r) \quad \forall p, q, r \in X\) (TRIANGLE INEQUALITY FAILS)
2. \(d(p + r, q + r) \neq d(p, q) \quad \forall p, q, r \in X\) (NOT TRANSLATION INVARIANT)
3. \(d(\alpha p, \alpha q) = |\alpha| d(p, q) \quad \forall p, q, \alpha \in \mathbb{R}\) (HOMOGENEOUS)
4. \(d\) does not induce a norm
5. balls in \((X, d)\) are in general NOT CONVEX

Proof:

1. Proof that the triangle inequality property fails to hold in \((X, d)\): Consider the following case\[11\]...

\[
d(p, r) \triangleq d((1, 0), (-0.5, 0)) = 0.767324 \ldots
\leq 0.756406 \ldots = 0.692330 \ldots + 0.064076 \ldots
= d((1, 0), (-0.5, 0.2)) + d((-0.5, 0.2), (-0.5, 0))
\triangleq d(p, q) + d(q, r)
\Longrightarrow \text{triangle inequality fails in } (X, d)
\]

2. Proof that \((X, d)\) is not translation invariant: Let \(r \triangleq (\frac{1}{2}, \frac{1}{2})\). Then...

\[
d(p + r, q + r) \triangleq d\left((\frac{1}{2}, \frac{1}{2}), \left(\frac{1}{2}, \frac{1}{2}\right)\right) = 0.229009 \ldots \neq \frac{1}{2}
\]

\[
= d\left((\frac{1}{2}, 0), \left(0, \frac{1}{2}\right)\right)
\triangleq d(p, q)
\Longrightarrow (X, d) \text{ is not translation invariant}
\]

3. Proof that \((X, d)\) is homogeneous:

Let \(r_{ap}\) be the magnitude of \(ap \triangleq (\alpha x_1, \alpha x_2, \ldots, \alpha x_n)\).

Let \(r_{aq}\) be the magnitude of \(aq \triangleq (\alpha y_1, \alpha y_2, \ldots, \alpha y_N)\).

Let \(\phi_a\) be the polar angle between \(ap\) and \(aq\).

---

\[11\] See experiment log file “lab_larc_distances.R2.txt” generated by the program “larc.exe”.

\[12\] See experiment log file “lab_larc_distances.R2.txt” generated by the program “larc.exe.”
(a) If \( r_p = 0 \) or \( r_q = 0 \) or \( \phi = 0 \) then \( d(p, q) \) is the Euclidean metric, which is homogeneous.

(b) lemmas:
\[
\begin{align*}
    r_{ap} &\triangleq \left( \sum_{n=1}^{N} \left| ax_n \right|^2 \right)^{\frac{1}{2}} = |a| \sum_{n=1}^{N} x_n^2 \triangleq |a|r_p \\
    r_{aq} &\triangleq \left( \sum_{n=1}^{N} \left| ay_n \right|^2 \right)^{\frac{1}{2}} = |a| \sum_{n=1}^{N} y_n^2 \triangleq |a|r_q \\
    \phi_q &\triangleq \text{arccos} \left( \frac{1}{r_{ap} r_{aq}} \sum_{n=1}^{N} [ax_n][ay_n] \right) = \text{arccos} \left( \frac{1}{r_{ap} r_{aq}} \sum_{n=1}^{N} x_n y_n \right) \triangleq \phi \\
    r(\theta; \alpha, aq) &\triangleq \left( \frac{r_{aq} - r_{ap}}{\phi} \right) \theta + r_{ap} = \left( \frac{r_{aq} - r_{ap}}{\phi} \right) \theta + ar_p = ar(\theta; p, q)
\end{align*}
\]

(c) If \( d(p, q) \) is not the Euclidean metric then …
\[
\pi d(p, aq) \triangleq R(p, aq) \quad \text{by definition of } d \quad \text{(Definition A.1 page 108)}
\]
\[
= \int_{0}^{\phi} \sqrt{[r(\theta; \alpha, aq)]^2 + \left[ \frac{dr(\theta; \alpha, aq)}{d\theta} \right]^2} \, d\theta \quad \text{by definition of } R \quad \text{(Definition A.1 page 108)}
\]
\[
= \int_{0}^{\phi} \sqrt{[ar(\theta; p, q)]^2 + \left[ \frac{d}{d\theta} r(\theta; p, q) \right]^2} \, d\theta \quad \text{by item (3b)}
\]
\[
= |a| \int_{0}^{\phi} \sqrt{[r(\theta; p, q)]^2 + \left[ \frac{d}{d\theta} r(\theta; p, q) \right]^2} \, d\theta \quad \text{by linearity of } \int_{0}^{\phi} \, d\theta \text{ operator}
\]
\[
\triangleq |a| R(p, q) \quad \text{by definition of } R \quad \text{(Definition A.1 page 108)}
\]

4. Proof that \( d \) does not induce a norm on \( X \): This follows directly from item (2) and Theorem D.5 (page 152).

5. Proof that balls (Definition B.4 page 126) in \( d \) are in general not convex (Definition 1.29 page 11):
   This is demonstrated graphically in Figure A.2 (page 120) and Figure A.3 (page 122).
   For an algebraic demonstration, consider the following:
   
   (a) Let \( \mathcal{B}((0, 1), 1) \) be the unit ball in \((\mathbb{R}^2, d)\) centered at \((0, 1)\).
   
   (b) Let \( p \triangleq (-0.70, -1.12), q \triangleq (0.70, -1.12), r \triangleq (0, -1.12), \text{ and } \lambda = \frac{1}{2} \).
   
   (c) Then \( d((0, 1), p) = 0.959536 \cdots < 1 \quad \Rightarrow \quad p \in \mathcal{B}((0, 1), 1) \text{ and} \)
   \( d((0, 1), q) = 0.959536 \cdots < 1 \quad \Rightarrow \quad q \in \mathcal{B}((0, 1), 1) \text{ but} \)
   \( d((0, 1), r) = 1.060688 \cdots > 1 \quad \Rightarrow \quad r \notin \mathcal{B}((0, 1), 1) \).
   
   (d) This implies that the set \( \mathcal{B}((0, 1), 1) \) is not convex because
   \[
   \lambda p + (1 - \lambda)q \triangleq \left( \frac{1}{2} (0.70, -1.12) + \left( 1 - \frac{1}{2} \right) (-0.70, -1.12) \right) \text{ by item (5b)}
   \]
   \[
   = (0, -1.12) \quad \triangleq r \quad \text{by item (5b)}
   \]
   \[
   \notin \mathcal{B}((0, 1), 1) \quad \text{by item (5c)}
   \]
   \[
   \Rightarrow \text{ the set } \mathcal{B}((0, 1), 1) \text{ is not convex} \quad \text{by Definition 1.29 page 11}
   \]

Remark A.1 (Lagrange arc distance versus Euclidean metric). As is implied by the metric balls illustrated in Figure A.2 (page 120) and Figure A.3 (page 122), the Lagrange arc distance \( d \) and Euclidean metric \( p \) are similar in the sense that they often lead to the same results\(^{14}\) in determining which of the two points \( q_1 \) or \( q_2 \) is “closer” to a point \( p \). But in some cases the two metrics lead to two different

\(^{13}\) See experiment log file “lab_larc_distances_R2.xtg” generated by the program “larc.exe”.

\(^{14}\) For empirical evidence of this, see Greenhoe (2016b).
\[
d((0, 1), (1, 0)) = \frac{\sqrt{2}}{2}
\]
\[
d((0, 1), (-1, 0)) = \frac{\sqrt{2}}{2}
\]
\[
d((0, 1), (0, -1)) = 1
\]
\[
d((1, 0), (0, -1)) = \frac{\sqrt{2}}{2}
\]
\[
d((1, 0), (-1, 0)) = 1
\]
\[
d((-1, 0), (0, -1)) = \frac{\sqrt{2}}{2}
\]
\[
d((0, 1), (2, 0)) = 0.8167968 \ldots
\]
\[
d((0, 1), (0, -2)) = 1.5346486 \ldots
\]
\[
d((0, 1), (-2, 1)) = 0.6966032 \ldots
\]

Figure A.1: Lagrange arc distance examples in \( \mathbb{R}^2 \)

\[
d((p, q_1) \triangleq d((1, 0), (-0.5, 0)) = 0.767324 \ldots
\]
\[
d((p, q_2) \triangleq d((1, 0), (-0.5, 0.75)) = 0.654039 \ldots
\]
\[
p ((p, q_1) \triangleq p ((1, 0), (-0.5, 0.5)) = 1.5
\]
\[
p ((p, q_2) \triangleq p ((1, 0), (-0.5, 0.75)) = \sqrt{(1.5)^2 + (0.75)^2} = 1.677050 \ldots
\]

That is, \( q_2 \) is closer than \( q_1 \) to \( p \) with respect to the Lagrange arc distance,
but \( q_1 \) is closer than \( q_2 \) to \( p \) with respect to the Euclidean metric.

A.5 Examples

Example A.2 (Lagrange arc distance in \( \mathbb{R}^2 \)). Figure A.1 (page 120) illustrates the Lagrange arc distance on some pairs of points in \( \mathbb{R}^2 \).

Example A.3 (Lagrange arc distance in \( \mathbb{R}^3 \)). Some examples of Lagrange arc distances in \( \mathbb{R}^3 \) are given in Table A.1 (page 121).
| \(d(( 0, 1, 0), ( 1, 0, 0))\) | \(\phi = \frac{\pi}{2}\) | 90° |
| \(d(( 0, 1, 0), ( 0, 0, 1))\) | \(\phi = \frac{\pi}{2}\) | 90° |
| \(d(( 0, 1, 0), ( 0, 0, -1))\) | \(\phi = \frac{\pi}{2}\) | 90° |
| \(d(( 0, 1, 0), ( -1, 0, 0))\) | \(\phi = \frac{\pi}{2}\) | 90° |
| \(d(( 0, 1, 0), ( 0, -1, 0))\) | \(\phi = \pi\) | 180° |
| \(d(( 1, 0, 0), ( 0, 0, 0))\) | \(\phi = \frac{\pi}{2}\) | 90° |
| \(d(( 1, 0, 0), ( 0, 0, -1))\) | \(\phi = \pi\) | 180° |
| \(d(( 1, 0, 0), ( -1, 0, 0))\) | \(\phi = \frac{\pi}{2}\) | 90° |
| \(d(( 0, 0, 1), ( 0, 0, 0))\) | \(\phi = \frac{\pi}{2}\) | 90° |
| \(d(( 0, 0, 1), ( -1, 0, 0))\) | \(\phi = \frac{\pi}{2}\) | 90° |
| \(d(( 0, 0, -1), ( 0, 0, 0))\) | \(\phi = \frac{\pi}{2}\) | 90° |
| \(d(( 0, -1, 0), ( 0, -1, 0))\) | \(\phi = \frac{\pi}{2}\) | 90° |
| \(d(( -1, 0, 0), ( 0, 0, 0))\) | \(\phi = \frac{\pi}{2}\) | 90° |
| \(d(( 0, 1, 0), ( 2, 0, 0))\) | \(\phi = \frac{\pi}{2}\) | 90° |
| \(d(( 0, 1, 0), ( 0, -2, 0))\) | \(\phi = \pi\) | 180° |
| \(d(( 0, 1, 0), ( -2, 1, 0))\) | \(\phi \approx 1.107\) | 63° |
| \(d(( 0, 1, 0), ( -1, 0, 1))\) | \(\phi = \pi\) | 90° |
| \(d(( 1, 1, 1), ( -\frac{\pi}{2}, \frac{\pi}{4}, -2))\) | \(\phi \approx 2.2466\) | 128.72° |

Table A.1: Some examples of Lagrange arc distances in \(\mathbb{R}^3\) (see Example A.3 page 120)

**Example A.4** (Lagrange arc distance balls in \(\mathbb{R}^2\)). Some unit balls in \(\mathbb{R}^2\) in with respect to the Lagrange arc distance are illustrated in Figure A.2 (page 120).

**Example A.5** (Lagrange arc distance balls in \(\mathbb{R}^3\)). Some unit balls in \(\mathbb{R}^3\) with respect to the Lagrange arc distance are illustrated in Figure A.3 (page 122).
centered at $(0,0,0)$

centered at $(0,0,-1)$

centered at $(0,0,-2)$

centered at $(0,0,-3)$

centered at $(0,0,-5)$

centered at $(0,0,-10)$

Figure A.3: unit Lagrange arc distance balls in $\mathbb{R}^3$
APPENDIX B
DISTANCE SPACES

“Tant que l’Algèbre et la Géométrie ont été séparées, leurs progrès ont été lents et leurs usages bornés; mais lorsque ces deux sciences se sont réunies, elles vers la perfection.”

Joseph-Louis Lagrange (1736–1813, Italian-French mathematician and astronomer  

“As long as algebra and geometry have been separated, their progress have been slow and their uses limited; but when these two sciences have been united, they have lent each other mutual forces, and have marched together with a rapid step towards perfection.”

B.1 Introduction and summary

Metric spaces provide a framework for analysis and have several very useful properties. Many of these properties follow in part from the triangle inequality. However, there are several applications in which the triangle inequality does not hold but in which we would still like to perform analysis. So the questions that naturally follow are:

1. What happens if we remove the triangle inequality all together?
2. What happens if we replace the triangle inequality with a generalized relation?

A distance space is a metric space without the triangle inequality constraint. Section B.2 introduces distance spaces and demonstrates that some properties commonly associated with metric spaces also hold in any distance space:

1 quote:  Lagrange (1795), page 271
translation:  Grattan-Guinness (1990) page 254

D1. \( \emptyset \) and \( X \) are open \hspace{1cm} \text{(Theorem B.2 page 126)}

D2. the intersection of a finite number of open sets is open \hspace{1cm} \text{(Theorem B.2 page 126)}

D3. the union of an arbitrary number of open sets is open \hspace{1cm} \text{(Theorem B.2 page 126)}

D4. every Cauchy sequence is bounded \hspace{1cm} \text{(Proposition B.1 page 129)}

D5. any subsequence of a Cauchy sequence is also Cauchy \hspace{1cm} \text{(Proposition B.2 page 129)}

D6. the Cantor Intersection Theorem holds \hspace{1cm} \text{(Theorem B.5 page 130)}

The following five properties (M1–M5) do hold in any metric space. However, the examples from Section B.2 listed below demonstrate that the five properties do not hold in all distance spaces:

M1. the metric function is continuous \hspace{1cm} \text{fails to hold in Example B.1–Example B.3}

M2. open balls are open \hspace{1cm} \text{fails to hold in Example B.1 and Example B.2}

M3. the open balls form a base for a topology \hspace{1cm} \text{fails to hold in Example B.1}

M4. the limits of convergent sequences are unique \hspace{1cm} \text{fails to hold in Example B.2}

M5. convergent sequences are Cauchy \hspace{1cm} \text{fails to hold in Example B.2}

Hence, Section B.2 answers question Q1.

APPENDIX C begins to answer question Q2 by first introducing a new function, called the power triangle function in a distance space \((X, d)\), as \(\tau(p, \sigma; x, y, z; d) \triangleq 2^p \left[ \frac{1}{2} d^p(x, z) + \frac{1}{2} d^p(z, y) \right]^{\frac{1}{p}}\) for some \((p, \sigma) \in \mathbb{R}^* \times \mathbb{R}\). APPENDIX C then goes on to use this function to define a new relation, called the power triangle inequality in \((X, d)\), and defined as

\[\Theta(p, \sigma; d) \triangleq \left\{ (x, y, z) \in X^3 \mid d(x, y) \leq \tau(p, \sigma; x, y, z; d) \right\} .\]

The power triangle inequality is a generalized form of the triangle inequality in the sense that the two inequalities coincide at \((p, \sigma) = (1, 1)\). Other special values include \((1, \sigma)\) yielding the relaxed triangle inequality (and its associated near metric space) and \((\infty, \sigma)\) yielding the \(\sigma\)-inframetric inequality (and its associated \(\sigma\)-inframetric space). Collectively, a distance space with a power triangle inequality is herein called a power distance space and denoted \((X, d, p, \sigma)\).\(^3\)

The power triangle function, at \(\sigma = \frac{1}{2}\), is a special case of the power mean with \(N = 2\) and \(\lambda_1 = \lambda_2 = \frac{1}{2}\). Power means have the elegant properties of being continuous and monontone with respect to a free parameter \(p\). From this it is easy to show that the power triangle function is also continuous and monontone with respect to both \(p\) and \(\sigma\). Special values of \(p\) yield operators coinciding with maximum, minimum, mean square, arithmetic mean, geometric mean, and harmonic mean. Power means are briefly described in APPENDIX D.3.3.\(^4\)

Section C.2 investigates the properties of power distance spaces. In particular, it shows for what values of \((p, \sigma)\) the properties M1–M5 hold. Here is a summary of the results in a power distance space \((X, d, p, \sigma)\), for all \(x, y, z \in X\):

\begin{align*}
(M1) \quad & \text{holds for any } (p, \sigma) \in (\mathbb{R}^* \setminus \{0\}) \times \mathbb{R}^+ \text{ such that } 2^\sigma = 2^{\frac{1}{2}} \quad \text{(Theorem C.5 page 142)} \\
(M2) \quad & \text{holds for any } (p, \sigma) \in (\mathbb{R}^* \setminus \{0\}) \times \mathbb{R}^+ \text{ such that } 2^\sigma \leq 2^{\frac{1}{2}} \quad \text{(Corollary C.6 page 140)} \\
(M3) \quad & \text{holds for any } (p, \sigma) \in (\mathbb{R}^* \setminus \{0\}) \times \mathbb{R}^+ \text{ such that } 2^\sigma \leq 2^{\frac{1}{2}} \quad \text{(Corollary C.5 page 139)} \\
(M4) \quad & \text{holds for any } (p, \sigma) \in \mathbb{R}^+ \times \mathbb{R}^+ \quad \text{(Theorem C.6 page 142)} \\
(M5) \quad & \text{holds for any } (p, \sigma) \in \mathbb{R}^+ \times \mathbb{R}^+ \quad \text{(Theorem C.3 page 141)}
\end{align*}

APPENDIX D.4 briefly introduces topological spaces. The open balls of any metric space form a base for a topology. This is largely due to the fact that in a metric space, open balls are open. Because of this, in metric spaces it is convenient to use topological structure to define and exploit analytic concepts such as continuity, convergence, closed sets, closure, interior, and accumulation point. For example, in a metric space, the traditional definition of defining continuity using open balls and

\(^3\) power triangle inequality: Definition C.3 page 136; power distance space: Definition C.2 page 135; examples of power distance space: Definition C.4 page 136;

\(^4\) power triangle function: Definition C.1 (page 135); power mean: Definition D.15 (page 164); power mean is continuous and monontone: Theorem D.14 (page 164); power triangle function is continuous and monontone: Corollary C.1 (page 136); Special values of \(p\): Corollary C.2 (page 136), Corollary D.2 (page 167)
the topological definition using open sets, coincide with each other. Again, this is largely because the open balls of a metric space are open.\(^5\)

However, this is not the case for all distance spaces. In general, the open balls of a distance space are not open, and they are not a base for a topology. In fact, the open balls of a distance space are a base for a topology if and only if the open balls are open. While the open sets in a distance space do induce a topology, it’s open balls may not.\(^6\)

A distance space (Definition B.1 page 125) can be defined as a metric space (Definition D.7 page 153) without the triangle inequality constraint. Much of the material in this section about distance spaces is standard in metric spaces. However, this paper works through this material again to demonstrate “how far we can go”, and can’t go, without the triangle inequality.

## B.2 Fundamental structure of distance spaces

### B.2.1 Definitions

**Definition B.1.**\(^7\) A function \(d \in \text{set } \mathbb{R}^{X \times X}\) (Definition 1.6 page 6) is a distance if

1. \(d(x, y) \geq 0 \quad \forall x, y \in X\) (non-negative) and
2. \(d(x, y) = 0 \iff x = y \quad \forall x, y \in X\) (non-degenerate) and
3. \(d(x, y) = d(y, x) \quad \forall x, y \in X\) (symmetric).

The pair \((X, d)\) is a distance space if \(d\) is a distance on a set \(X\).

**Definition B.2.**\(^8\) Let \((X, d)\) be a distance space and \(2^X\) be the power set of \(X\) (Definition 1.8 page 6). The diameter in \((X, d)\) of a set \(A \subseteq 2^X\) is

\[
\text{diam } A \triangleq \begin{cases} 
0 & \text{for } A = \emptyset \\
\sup \{d(x, y) \mid x, y \in A\} & \text{otherwise}
\end{cases}
\]

**Definition B.3.**\(^9\) Let \((X, d)\) be a distance space. Let \(2^X\) be the power set (Definition 1.8 page 6) of \(X\). A set \(A\) is bounded in \((X, d)\) if \(A \subseteq 2^X\) and \(\text{diam } A < \infty\).

### B.2.2 Properties

**Theorem B.1.**\(^10\) Let \((x_n)_{n \in \mathbb{Z}}\) be a sequence in a distance space \((X, d)\). The distance space \((X, d)\) does not necessarily have all the nice properties that a metric space (Definition D.7 page 153) has. In particular, note the following:

\(^5\)open ball: Definition B.4 page 126; metric space: Definition D.7 page 153; base: Definition D.17 page 168; topology: Definition D.16 page 168; open: Definition B.5 page 126; continuity in topological space: Definition D.19 page 169; convergence in distance space: Definition B.7 page 128; convergence in topological space: Definition D.20 page 170; closed set: Definition D.16 page 168; closure, interior, accumulation point: Definition D.18 page 169; coincidence in all metric spaces and some power distance spaces: Theorem C.2 page 140;

\(^6\)if and only if statement: Theorem B.3 page 128; open sets of a distance space induce a topology: Corollary B.1 page 127;

\(^7\)Menger (1928) page 76 (“Abstand a b definiert ist...” (distance from a to b is defined as...”)), \(\bowtie\) Wilson (1931b) page 361 (§1., “distance”, “semi-metric space”), \(\bowtie\) Blumenthal (1938) page 38, \(\bowtie\) Blumenthal (1953) page 7 (“DEFINITION 5.1. A distance space is called semimetric provided...”), \(\bowtie\) Galvin and Shore (1984) page 67 (“distance function”), \(\bowtie\) Laos (1998) page 118 (“distance space”), \(\bowtie\) Khamri and Kirk (2001) page 13 (“semimetric space”), \(\bowtie\) Bessenyei and Pales (2014) page 2 (“semimetric space”), \(\bowtie\) Deza and Deza (2014) page 3 (“distance (or dissimilarity)”)\)

\(^8\)in metric space: \(\bowtie\) Hausdorff (1937), page 166, \(\bowtie\) Copson (1968), page 23, \(\bowtie\) Michel and Herget (1993), page 267, \(\bowtie\) Molchanov (2005) page 389

\(^9\)in metric space: \(\bowtie\) Thron (1966), page 154 (definition 19.5), \(\bowtie\) Bruckner et al. (1997) page 356

\(^10\)\(\bowtie\) Greenhoe (2016a)
1. \( d \) is a distance in \((X, d)\) \(\implies\) \( d \) is continuous in \((X, d)\) (Example B.3 page 133).
2. \( B \) is an open ball in \((X, d)\) \(\implies\) \( B \) is open in \((X, d)\) (Example B.2 page 132).
3. \( B \) is the set of all open balls in \((X, d)\) \(\implies\) \( B \) is a base for a topology on \(X\) (Example B.2 page 132).\(^{11}\)
4. \( \{x_n\} \) is convergent in \((X, d)\) \(\implies\) limit is unique (Example B.1 page 131).
5. \( \{x_n\} \) is convergent in \((X, d)\) \(\implies\) \( \{x_n\} \) is Cauchy in \((X, d)\) (Example B.2 page 132).

### B.3 Open sets in distance spaces

#### B.3.1 Definitions

**Definition B.4.** \(^{12}\) Let \((X, d)\) be a distance space (Definition B.1 page 125).
An open ball centered at \(x\) with radius \(r\) is the set \(B(x, r) \triangleq \{y \in X \mid d(x, y) < r\}\).
A closed ball centered at \(x\) with radius \(r\) is the set \(\overline{B}(x, r) \triangleq \{y \in X \mid d(x, y) \leq r\}\).

**Definition B.5.** Let \((X, d)\) be a distance space. Let \(X \backslash A\) be the set difference of \(X\) and a set \(A\). A set \(U\) is open in \((X, d)\) if \(U \in 2^X\) and for every \(x\) in \(U\) there exists \(r \in \mathbb{R}^+\) such that \(B(x, r) \subseteq U\). A set \(U\) is an open set in \((X, d)\) if \(U\) is open in \((X, d)\). A set \(D\) is closed in \((X, d)\) if \(D\) is closed in \((X, d)\).

#### B.3.2 Properties

**Theorem B.2.** \(^{13}\) Let \((X, d)\) be a distance space. Let \(N\) be any (finite) positive integer. Let \(\Gamma\) be a set possibly with an uncountable number of elements.

1. \(X\) is open.
2. \(\emptyset\) is open.
3. each element in \(\{U_n\}_{n=1,2,\ldots,N}\) is open \(\implies\) \(\bigcap_{n=1}^{N} U_n\) is open.
4. each element in \(\{U_\gamma \subseteq 2^X \mid \gamma \in \Gamma\}\) is open \(\implies\) \(\bigcup_{\gamma \in \Gamma} U_\gamma\) is open.

\(^{11}\) Heath (1961) page 810 (Theorem), \(^{12}\) Aliprantis and Burkinshaw (1998), page 35
\(^{13}\) in metric space: \(^{13}\) Aliprantis and Burkinshaw (1998), page 35
3. Proof that $\bigcup U_r$ is open in $(X,d)$:
   (a) By definition of open set (Definition B.5 page 126), $\bigcup U_r$ is open $\iff \forall x \in \bigcup U_r \exists r$ such that $B(x,r) \subseteq \bigcup U_r$.
   (b) If $x \in \bigcup U_r$, then there is at least one $U \in \bigcup U_r$ that contains $x$.
   (c) By the left hypothesis in (4), that set $U$ is open and so for that $x$, $\exists r$ such that $B(x,r) \subseteq U \subseteq \bigcup U_r$.
   (d) Therefore, $\bigcup U_r$ is open in $(X,d)$.

4. Proof that $U_1$ and $U_2$ are open $\implies U_1 \cap U_2$ is open:
   (a) By definition of open set (Definition B.5 page 126), $U_1 \cap U_2$ is open $\iff \forall x \in U_1 \cap U_2 \exists r$ such that $B(x,r) \subseteq U_1 \cap U_2$.
   (b) By the left hypothesis above, $U_1$ and $U_2$ are open; and by the definition of open sets (Definition B.5 page 126), there exists $r_1$ and $r_2$ such that $B(x,r_1) \subseteq U_1$ and $B(x,r_2) \subseteq U_2$.
   (c) Let $r = \min \{r_1, r_2\}$. Then $B(x,r) \subseteq U_1$ and $B(x,r) \subseteq U_2$.
   (d) By definition of set intersection $\cap$ then, $B(x,r) \subseteq U_1 \cap U_2$.
   (e) By definition of open set (Definition B.5 page 126), $U_1 \cap U_2$ is open.

5. Proof that $\bigcap_{n=1}^{N} U_n$ is open (by induction):
   (a) Proof for $N = 1$ case: $\bigcap_{n=1}^{N} U_n = \bigcap_{n=1}^{1} U_n = U_1$ is open by hypothesis.
   (b) Proof that $N$ case $\implies N + 1$ case:
      
      \[
      \bigcap_{n=1}^{N+1} U_n = \left( \bigcap_{n=1}^{N} U_n \right) \cap U_{N+1} \quad \text{by property of } \cap
      \]
      
      \[
      \implies \text{open} \quad \text{by "} N \text{ case" hypothesis and (4) lemma page 127}
      \]

\[\square\]

Corollary B.1. Let $(X,d)$ be a distance space. The set $T \triangleq \{ U \in 2^X | U \text{ is open in } (X,d) \}$ is a topology on $X$, and $(X,T)$ is a topological space.

\[\square\]

\(\text{Proof:}\) This follows directly from the definition of an open set (Definition B.5 page 126), Theorem B.2 (page 126), and the definition of topology (Definition D.16 page 168).

Of course it is possible to define a very large number of topologies even on a finite set with just a handful of elements,\(^\text{14}\) and it is possible to define an infinite number of topologies even on a linearly ordered infinite set like the real line $(\mathbb{R}, \leq)$.\(^\text{15}\) Be that as it may, Definition B.6 (next definition) defines a single but convenient topological space in terms of a distance space. Note that every metric space conveniently and naturally induces a topological space because the open balls of the metric space form a base for the topology. This is not the case for all distance spaces. But if the open balls of a distance space are all open, then those open balls induce a topology (next theorem)\(^\text{16}\).

\(^\text{14}\)For a finite set $X$ with $n$ elements, there are 29 topologies on $X$ if $n = 3$, 6942 topologies on $X$ if $n = 5$, and and 8,977,053,873,043 (almost 9 trillion) topologies on $X$ if $n = 10$. References: \(\square\) Sloane (2014) (http://oeis.org/A000798), \(\square\) Brown and Watson (1996), page 31, \(\square\) Comtet (1974) page 229, \(\square\) Comtet (1966), \(\square\) Chatterji (1967), page 7, \(\square\) Evans et al. (1967), \(\square\) Krishnamurthy (1966), page 157

\(^\text{15}\)For examples of topologies on the real line, see the following: \(\square\) Adams and Franzosa (2008) page 31 ("six topologies on the real line"), \(\square\) Salzmann et al. (2007) pages 64–70 (Weird topologies on the real line), \(\square\) Murdeshwar (1990) page 53 ("often used topologies on the real line"), \(\square\) Joshi (1983) pages 85–91 (§4.2 Examples of Topological Spaces)

\(^\text{16}\)metric space: Definition D.7 page 153; open ball: Definition B.4 page 126; base: Definition D.17 page 168; topology: Definition D.16 page 168; not all open balls are open in a distance space: Example B.1 (page 131) and Example B.2 (page 132);
**Definition B.6.** Let \((X, d)\) be a distance space. The set \(T \triangleq \{ U \in 2^X | U \text{ is open in } (X, d) \}\) is the topology induced by \((X, d)\) on \(X\). The pair \((X, T)\) is called the topological space induced by \((X, d)\).

For any distance space \((X, d)\), no matter how strange, there is guaranteed to be at least one topological space induced by \((X, d)\)—and that is the indiscrete topological space (Example D.12 page 168) because for any distance space \((X, d)\), \(\emptyset\) and \(X\) are open sets in \((X, d)\) (Theorem B.2 page 126).

**Theorem B.3.** Let \(B\) be the set of all open balls in a distance space \((X, d)\).

\[
\{\text{every open ball in } B \text{ is open} \} \iff \{ \text{B is a base for a topology} \}
\]

\(\circ\) Proof:

\[
\text{every open ball in } B \text{ is open}
\]

\[
\implies \text{for every } x \in B, \text{ there exists } r \in \mathbb{R}^+ \text{ such that } B(x, r) \subseteq B_y \text{ by definition of open (Definition B.5 page 126)}
\]

\[
\implies \{ \text{for every } x \in X \text{ and for every } B_y \in B \text{ containing } x, \text{ there exists } B_x \in B \text{ such that } x \subseteq B_y \text{ } \}
\]

\[
\implies B \text{ is a base for } T \text{ by Theorem D.16 page 168}
\]

\[
\implies \{ \text{for every } x \in X \text{ and for every } U \subseteq T \text{ containing } x, \text{ there exists } B_x \in B \text{ such that } x \subseteq B_x \subseteq U \text{ by Theorem D.16 page 168} \}
\]

\[
\implies \{ \text{for every } x \in X \text{ and for every } B_y \in B \subseteq T \text{ containing } x, \text{ there exists } B_x \in B \text{ such that } x \subseteq B_x \subseteq B_y \text{ by definition of base (Definition D.17 page 168)} \}
\]

\[
\implies \{ \text{there exists } B_x \in B \text{ such that } x \subseteq B_x \subseteq B_y \}
\]

\[
\implies \text{every open ball in } B \text{ is open by definition of open (Definition B.5 page 126)}
\]

\(\square\)

## B.4 Sequences in distance spaces

### B.4.1 Definitions

**Definition B.7.** \(^{17}\) Let \((x_n \in X)_{n \in \mathbb{Z}}\) be a sequence in a distance space \((X, d)\). The sequence \((x_n)\) converges to a limit \(x\) if for any \(\varepsilon \in \mathbb{R}^+\), there exists \(N \in \mathbb{Z}\) such that for all \(n > N\), \(d(x_n, x) < \varepsilon\).

This condition can be expressed in any of the following forms:

1. The limit of the sequence \((x_n)\) is \(x\).
2. The sequence \((x_n)\) is convergent with limit \(x\).
3. \(\lim_{n \to \infty} (x_n) = x\).
4. \((x_n) \to x\).

A sequence that converges is **convergent**.

**Definition B.8.** \(^{18}\) Let \((x_n \in X)_{n \in \mathbb{Z}}\) be a sequence in a distance space \((X, d)\). The sequence \((x_n)\) is a **Cauchy sequence** in \((X, d)\) if for every \(\varepsilon \in \mathbb{R}^+\), there exists \(N \in \mathbb{Z}\) such that for all \(n, m > N\), \(d(x_n, x_m) < \varepsilon\) (Cauchy condition).

**Definition B.9.** \(^{19}\) Let \((x_n \in X)_{n \in \mathbb{Z}}\) be a sequence in a distance space \((X, d)\). The sequence \((x_n \in X)_{n \in \mathbb{Z}}\) is **complete** in \((X, d)\) if \((x_n) \text{ is CAUCHY in } (X, d) \implies (x_n) \text{ is CONVERGENT in } (X, d)\).

\(^{17}\) in metric space: \(\bowtie\) Rosenlicht (1968) page 45, \(\bowtie\) Giles (1987) page 37 (3.2 Definition), \(\bowtie\) Khamsi and Kirk (2001) page 13 (Definition 2.1) “→” symbol: \(\bowtie\) Leatham (1995) page 13 (section III.11)

\(^{18}\) in metric space: \(\bowtie\) Apostol (1975) page 73 (4.7), \(\bowtie\) Rosenlicht (1968) page 51

\(^{19}\) in metric space: \(\bowtie\) Rosenlicht (1968) page 52
B.4.2 Properties

Proposition B.1. \(^{20}\) Let \(\{x_n \in X\}_{n \in \mathbb{Z}}\) be a sequence in a distance space \((X, d)\).
\[
\{ (x_n) \text{ is Cauchy in } (X, d) \} \quad \implies \quad \{ (x_n) \text{ is bounded in } (X, d) \}
\]
\text{Proof:}

\[(x_n) \text{ is Cauchy } \implies \text{ for every } \varepsilon \in \mathbb{R}^+, \exists N \in \mathbb{Z} \text{ such that } \forall n, m > N, d(x_n, x_m) < \varepsilon \quad (\text{Definition B.8 page 128})
\]
\[
\implies \exists N \in \mathbb{Z} \text{ such that } \forall n, m > N, d(x_n, x_m) < 1 \quad \text{(arbitrarily choose } \varepsilon \triangleq 1)\]
\[
\implies \exists N \in \mathbb{Z} \text{ such that } \forall n, m \in \mathbb{Z}, d(x_n, x_{m+1}) < \max \{ \{1 \} \cup \{ d(x_p, x_q) | p, q \neq N \} \}
\]
\[
\implies (x_n) \text{ is bounded} \quad \text{(by Definition B.3 page 125)}
\]

Proposition B.2. \(^{21}\) Let \(\{x_n \in X\}_{n \in \mathbb{Z}}\) be a sequence in a distance space \((X, d)\). Let \(f \in \mathbb{Z}^Z\) (Definition 1.6 page 6) be a strictly monotone function such that \(f(n) < f(n + 1)\).
\[
\{ (x_n)_{n \in \mathbb{Z}} \text{ is Cauchy} \} \implies \{ (f(x_n))_{n \in \mathbb{Z}} \text{ is Cauchy} \}
\]
\text{Proof:}

\[(x_n)_{n \in \mathbb{Z}} \text{ is Cauchy } \implies \text{ for any given } \varepsilon > 0, \exists N \text{ such that } \forall n, m > N, d(x_n, x_m) < \varepsilon \quad \text{(by Definition B.8 page 128)}
\]
\[
\implies \text{ for any given } \varepsilon > 0, \exists N' \text{ such that } \forall f(n), f(m) > N', d(f(x_n), f(x_m)) < \varepsilon 
\]
\[
\implies (f(x_n))_{n \in \mathbb{Z}} \text{ is Cauchy} \quad \text{(by Definition B.8 page 128)}
\]

Theorem B.4. \(^{22}\) Let \((X, d)\) be a distance space. Let \(A^-\) be the closure (Definition D.18 page 169) of a \(A\) in a topological space induced by \((X, d)\).
\[
\{\begin{array}{l}
1. \text{ Limits are unique in } (X, d) \quad (\text{Definition B.7 page 128}) \\
2. (A, d) \text{ is complete in } (X, d) \quad (\text{Definition B.9 page 128})
\end{array}\}
\implies A \text{ is closed in } (X, d)
\]
\text{Proof:}

1. Proof that \(A \subseteq A^-\): by Lemma D.5 page 169

2. Proof that \(A^- \subseteq A\) (proof that \(x \in A^- \implies x \in A\)):
   (a) Let \(x\) be a point in \(A^-\) (\(x \in A^-\)).
   (b) Define a sequence of open balls \(\{B(x, \frac{1}{n}) \cup B(x, \frac{1}{n}) \cup B(x, \frac{1}{n}) \cup \ldots\}\).
   (c) Define a sequence of points \(x_1, x_2, x_3, \ldots\) such that \(x_n \in B(x, \frac{1}{n}) \cap A\).
   (d) Then \((x_n)\) is convergent in \(X\) with limit \(x\) by Definition B.7 page 128
   (e) and \((x_n)\) is Cauchy in \(A\) by Definition B.8 page 128.
   (f) By the hypothesis 2, \((x_n)\) is therefore also convergent in \(A\).
   Let this limit be \(y\). Note that \(y \in A\).

---

\(^{20}\) in metric space: \(\equiv\) Giles (1987) page 49 (Theorem 3.30)

\(^{21}\) in metric space: \(\equiv\) Rosenlicht (1968) page 52

\(^{22}\) in metric space: \(\equiv\) Kubrusly (2001) page 128 (Theorem 3.40), Haaser and Sullivan (1991) page 75 (6-10, 6-11 Propositions), Bryant (1985) page 40 (Theorem 3.6, 3.7), Sutherland (1975) pages 123–124
(g) By hypothesis 1, limits are unique, so y = x.
(h) Because y ∈ A (item (2f)) and y = x (item (2g)), so x ∈ A.
(i) Therefore, x ∈ A⁻ → x ∈ A and A⁻ ⊆ A.

Proposition B.3. 23 Let \( (x_n)_{n \in \mathbb{Z}} \) be a sequence in a distance space \((X, d)\). Let \( f : \mathbb{Z} \to \mathbb{Z} \) be a strictly increasing function such that \( f(n) < f(n + 1) \).

\[
\begin{align*}
\lim_{n \to \infty} x_n & \implies \forall \epsilon > 0, \exists N \text{ such that } \forall n > N, d(x_n, x) < \epsilon \\
\implies & \forall \epsilon > 0, \exists f(N) \text{ such that } \forall f(n) > f(N), d(x_{f(n)}, x) < \epsilon \\
\implies & \lim_{n \to \infty} x_{f(n)} = x
\end{align*}
\]

\( \therefore \text{sequence converges to limit } x \implies \text{subsequence converges to the same limit } x \)

\( \Box \)

Theorem B.5 (Cantor intersection theorem). 24 Let \((X, d)\) be a distance space (Definition B.1 page 125), \((A_n)_{n \in \mathbb{Z}}\) a sequence with each \(A_n \subseteq 2^X\), and \(|A|\) the number of elements in A.

1. \((X, d)\) is complete \( \iff \lim_{n \to \infty} A_n \neq \emptyset \)
2. \(A_n\) is closed \( \iff \lim_{n \to \infty} A_n \neq \emptyset \)
3. \(\text{diam } A_n \geq \text{diam } A_{n+1}\)
4. \(\exists n \in \mathbb{N} \text{ such that } \text{diam } A_n < \epsilon \)

\( \implies \left\{ \lim_{n \to \infty} A_n \right\} = 1 \)

\( \Box \)

Theorem B.5 (Cantor intersection theorem). 24 Let \((X, d)\) be a distance space (Definition B.1 page 125), \((A_n)_{n \in \mathbb{Z}}\) a sequence with each \(A_n \subseteq 2^X\), and \(|A|\) the number of elements in A.

1. \((X, d)\) is complete \( \iff \lim_{n \to \infty} A_n \neq \emptyset \)
2. \(A_n\) is closed \( \iff \lim_{n \to \infty} A_n \neq \emptyset \)
3. \(\text{diam } A_n \geq \text{diam } A_{n+1}\)
4. \(\exists n \in \mathbb{N} \text{ such that } \text{diam } A_n < \epsilon \)

\( \implies \left\{ \lim_{n \to \infty} A_n \right\} = 1 \)

\( \Box \)

THEOREM B.5 (CANTOR INTERSECTION THEOREM). 24 Let \((X, d)\) be a distance space (Definition B.1 page 125), \((A_n)_{n \in \mathbb{Z}}\) a sequence with each \(A_n \subseteq 2^X\), and \(|A|\) the number of elements in A.

1. \((X, d)\) is complete \( \iff \lim_{n \to \infty} A_n \neq \emptyset \)
2. \(A_n\) is closed \( \iff \lim_{n \to \infty} A_n \neq \emptyset \)
3. \(\text{diam } A_n \geq \text{diam } A_{n+1}\)
4. \(\exists n \in \mathbb{N} \text{ such that } \text{diam } A_n < \epsilon \)

\( \implies \left\{ \lim_{n \to \infty} A_n \right\} = 1 \)

\( \Box \)

Proof:

1. Proof that \(|\bigcap_{n \in \mathbb{Z}} A_n| < 2^\): 
   (a) Let \(A \neq \bigcap_{n \in \mathbb{N}} A_n\).
   (b) \(x \neq y\) and \((x, y) \in A \implies d(x, y) > 0\) and \(|x, y| \subseteq A\) \(\forall n\)
   (c) \(\exists n\) such that \(\text{diam } A_n < d(x, y)\) by left hypothesis 4
   (d) \(\implies \exists n\) such that \(\sup\{d(x, y) | x, y \in A_n\} < d(x, y)\)
   (e) This is a contradiction, so \(|x, y| \subseteq A\) and \(|\bigcap_{n \in \mathbb{N}} A_n| < 2^\).

2. Proof that \(|\bigcap_{n \in \mathbb{N}} A_n| > 0^\): 
   (a) Let \(x_n \in A_n \) and \(x_m \in A_m\)
   (b) \(\forall \epsilon, \exists N \in \mathbb{N} \text{ such that } A_N < \epsilon \)
   (c) \(\forall m, n > N, x_n \in A_n \subseteq A_N \) and \(x_m \in A_m \subseteq A_N\)
   (d) \(d(x_n, x_m) \leq \text{diam } A_N < \epsilon \) \(\implies \{x_n\} \text{ is a Cauchy sequence}\)
   (e) Because \(|x_n|\) is complete, \(x_n \to x\).
   (f) \(\implies x \in (A_n)^\sim = A_n\)
   (g) \(|A_n| > 0^\)

\( \Box \)

\(23\) in metric space; \(\Box\) Rosenlicht (1968) page 46
\(24\) in metric space; \(\Box\) Davis (2005), page 28, \(\Box\) Hausdorff (1937), page 150
Definition B.10. Let $(X, d)$ be a distance space. Let $C$ be the set of all convergent sequences in $(X, d)$. The distance function $d$ is **continuous** in $(X, d)$ if

$$
\langle x_n \rangle, \langle y_n \rangle \in C \implies \lim_{n \to \infty} d(x_n, y_n) = d \left( \lim_{n \to \infty} x_n, \lim_{n \to \infty} y_n \right).
$$

A distance function is **discontinuous** if it is not continuous.

Remark B.1. Rather than defining continuity of a distance function in terms of the sequential characterization of continuity as in Definition B.10 (previous), we could define continuity using an inverse image characterization of continuity” (Definition B.6 page 128). Assuming an equivalent topological space is used for both characterizations, the two characterizations are equivalent (Theorem D.20 page 171). In fact, one could construct an equivalence such as the following:

$$
\left\{ \begin{array}{l}
d \text{ is continuous in } \mathbb{R}^{X^2} \\
(\text{Definition D.19 page 169}) \end{array} \right\} \iff \left\{ \begin{array}{l}
\langle x_n \rangle, \langle y_n \rangle \in C \implies \lim_{n \to \infty} d(x_n, y_n) = d \left( \lim_{n \to \infty} x_n, \lim_{n \to \infty} y_n \right) \\
(\text{Definition D.20 page 170}) \end{array} \right\}
$$

Note that just as $\langle x_n \rangle$ is a sequence in $X$, so the ordered pair $\langle \langle x_n \rangle, \langle y_n \rangle \rangle$ is a sequence in $X^2$. The remainder follows from Theorem D.20 (page 171). However, use of the inverse image characterization is somewhat troublesome because we would need a topology on $X^2$, and we don’t immediately have one defined and ready to use. In fact, we don’t even immediately have a distance space on $X^2$ defined or even open balls in such a distance space. The result is, for the scope of this paper, it is arguably not worthwhile constructing the extra structure, but rather instead this paper uses the sequential characterization as a definition (as in Definition B.10).

B.5 Examples

Similar distance functions and several of the observations for the examples in this section can be found in Blumenthal (1953) pages 8–13.

In a metric space, all open balls are open, the open balls form a base for a topology, the limits of convergent sequences are unique, and the metric function is continuous. In the distance space of the next example, none of these properties hold.

Example B.1. Let $(x, y)$ be an ordered pair in $\mathbb{R}^2$. Let $(a : b)$ be an open interval and $(a : b]$ a half-open interval in $\mathbb{R}$. Let $|x|$ be the absolute value of $x \in \mathbb{R}$. The function $d(x, y) \in \mathbb{R}^{\mathbb{R} \times \mathbb{R}}$ such that

$$
d(x, y) \triangleq \begin{cases} 
y & \forall (x, y) \in \{4\} \times \{0:2\} \\
x & \forall (x, y) \in \{0:2\} \times \{4\} \\
|x - y| & \text{otherwise} \\
\end{cases}
$$

is a distance on $\mathbb{R}$.

Note some characteristics of the distance space $(\mathbb{R}, d)$:

1. $(\mathbb{R}, d)$ is not a metric space because $d$ does not satisfy the triangle inequality:

   $$
d(0,4) \triangleq |0 - 4| = 4 \not\leq 2 = |0 - 1| + 1 \not\triangleq d(0,1) + d(1,4)
$$

2. Not every open ball in $(\mathbb{R}, d)$ is open.

   For example, the open ball $B(3, 2)$ is not open because $4 \in B(3, 2)$ but for all $0 < \epsilon < 1$

   $$
   B(4, \epsilon) = (4 - \epsilon : 4 + \epsilon) \cup (0 : \epsilon) \not\subseteq (1 : 5) = B(3, 2)
   $$

\[25\] Blumenthal (1953) page 9 (Definition 6.3)

\[26\] A similar distance function $d$ and item (4) page 132 can in essence be found in Blumenthal (1953) page 8. Definitions for Example B.1: $(x, y)$: Definition 1.3 (page 5); $(a : b)$ and $(a : b]$: Definition 1.27 (page 11); $|x|$: Definition 1.28 (page 11); $\mathbb{R}^{\mathbb{R} \times \mathbb{R}}$: Definition 1.6 (page 6); distance: Definition 1.1 (page 125); open ball: Definition 1.4 (page 126); open: Definition 1.5 (page 126); base: Definition D.17 (page 168); topology: Definition D.16 (page 168); open set: Definition B.5 (page 126); topological space induced by $(\mathbb{R}, d)$: Definition B.6 (page 128); discontinuous: Definition B.10 (page 131);
3. The open balls of \((\mathbb{R}, d)\) do not form a base for a topology on \(\mathbb{R}\).
   This follows directly from item (2) and Theorem B.3 (page 128).

4. In the distance space \((\mathbb{R}, d)\), limits are not unique;
   For example, the sequence \(\langle \frac{1}{n} \rangle\) converges both to the limit 0 and the limit 4 in \((\mathbb{R}, d)\):
   \[
   \lim_{n \to \infty} d(x_n, 0) = \lim_{n \to \infty} (\frac{1}{n}, 0) = \lim_{n \to \infty} |\frac{1}{n} - 0| = 0 \quad \implies \langle \frac{1}{n} \rangle \to 0
   \]
   \[
   \lim_{n \to \infty} d(x_n, 4) = \lim_{n \to \infty} (\frac{1}{n}, 4) = \lim_{n \to \infty} |\frac{1}{n} - 4| = 0 \quad \implies \langle \frac{1}{n} \rangle \to 4
   \]

5. The topological space \((X, T)\) induced by \((\mathbb{R}, d)\) also yields limits of 0 and 4 for the sequence \(\langle \frac{1}{n} \rangle\), just as it does in item (4). This is largely due to the fact that, for small \( \varepsilon \), the open balls \( B(0, \varepsilon) \) and \( B(4, \varepsilon) \) are open.

   \[\text{B(0,} \varepsilon) \text{ is open } \implies \text{ for each} \ U \in T \text{ that contains 0, } \exists N \in \mathbb{N} \text{ such that } \frac{1}{n} \in U \quad \forall n > N\]
   \[\iff \langle \frac{1}{n} \rangle \to 0 \quad \text{by definition of convergence (Definition D.20 page 170)}\]

   \[\text{B(4,} \varepsilon) \text{ is open } \implies \text{ for each} \ U \in T \text{ that contains 4, } \exists N \in \mathbb{N} \text{ such that } \frac{1}{n} \in U \quad \forall n > N\]
   \[\iff \langle \frac{1}{n} \rangle \to 4 \quad \text{by definition of convergence (Definition D.20 page 170)}\]

6. The distance function \(d\) is discontinuous (Definition B.10 page 131):
   \[
   \lim_{n \to \infty} (d(1 - \frac{1}{n}, 4 - \frac{1}{n})) = \lim_{n \to \infty} \left( \left| (1 - \frac{1}{n}) - (4 - \frac{1}{n}) \right| \right) = |1 - 4| = 3 \neq 4 = d(0, 4)
   \]
   \[= d \left( \lim_{n \to \infty} \langle 1 - \frac{1}{n} \rangle , \lim_{n \to \infty} \langle 4 - \frac{1}{n} \rangle \right)\]

In a metric space, all convergent sequences are also Cauchy. However, this is not the case for all distance spaces, as demonstrated next:

Example B.2. 27 The function \(d(x, y) \in \mathbb{R}^{\mathbb{R} \times \mathbb{R}}\) such that
\[
d(x, y) \triangleq \begin{cases} 
|x - y| & \text{for } x = 0 \text{ or } y = 0 \text{ or } x = y \\
1 & \text{otherwise} 
\end{cases} \quad \text{(Euclidean)} \quad \text{(discrete)}
\]
is a distance on \(\mathbb{R}\).

Note some characteristics of the distance space \((\mathbb{R}, d)\):

1. \((X, d)\) is not a metric space because the triangle inequality does not hold:
   \[d \left( \frac{1}{2}, \frac{1}{2} \right) = 1 \neq \frac{1}{2} = \left| \frac{1}{2} - 0 \right| + \left| 0 - \frac{1}{2} \right| = d \left( \frac{1}{2}, 0 \right) + d \left( 0, \frac{1}{2} \right)\]

2. The open ball \(B \left( \frac{1}{2}, \frac{1}{2} \right)\) is not open because for any \( \varepsilon \in \mathbb{R}^+ \), no matter how small,
   \[B(0, \varepsilon) = (-\varepsilon : +\varepsilon) \not\subseteq \{0, \frac{1}{2}\} = \left\{ x \in \mathbb{R} | d \left( \frac{1}{2}, x \right) < \frac{1}{2} \right\} \triangleq B \left( \frac{1}{2}, \frac{1}{2} \right)\]

3. Even though not all the open balls are open, it is still possible to have an open set in \((X, d)\). For example, the set \(U \triangleq \{1, 2\}\) is open:
   \[B(1, 1) \triangleq \{ x \in \mathbb{R} | d(1, x) < 1 \} = \{1\} \subseteq \{1, 2\} \triangleq U\]
   \[B(2, 1) \triangleq \{ x \in \mathbb{R} | d(2, x) < 1 \} = \{2\} \subseteq \{1, 2\} \triangleq U\]

4. By item (2) and Theorem B.3 (page 128), the open balls of \((\mathbb{R}, d)\) do not form a base for a topology on \(\mathbb{R}\).

5. Even though the open balls in \((\mathbb{R}, d)\) do not induce a topology on \(X\), it is still possible to find a set of open sets in \((X, d)\) that is a topology. For example, the set \(\emptyset, \{1, 2\}, \mathbb{R}\) is a topology on \(\mathbb{R}\).

---

27The distance function \(d\) and item (7) page 133 can in essence be found in Blumenthal (1953) page 9
6. In \((\mathbb{R}, d)\), limits of \textit{convergent} sequences are \textit{unique}:

\[
(x_n) \to x \quad \implies \quad \lim_{n \to \infty} d(x_n, x) = \begin{cases} 
\lim |x_n - 0| = 0 & \text{for } x = 0 \\
|x - x| = 0 & \text{for constant } (x_n) \text{ for } n > N \\
1 \neq 0 & \text{otherwise}
\end{cases}
\]

which says that there are only two ways for a sequence to converge: either \(x = 0\) or the sequence eventually becomes constant (or both). Any other sequence will \textit{diverge}. Therefore we can say the following:

(a) If \(x = 0\) and the sequence is not constant, then the limit is \textit{unique} and 0.

(b) If \(x = 0\) and the sequence is constant, then the limit is \textit{unique} and 0.

(c) If \(x \neq 0\) and the sequence is constant, then the limit is \textit{unique} and \(x\).

(d) If \(x \neq 0\) and the sequence is not constant, then the sequence diverges and there is no limit.

7. In \((\mathbb{R}, d)\), a \textit{convergent} sequence is not necessarily \textit{Cauchy}. For example,

(a) the sequence \((\frac{1}{n})_{n \in \mathbb{N}}\) is \textit{convergent} with limit 0: \(\lim_{n \to \infty} d(\frac{1}{n}, 0) = \lim_{n \to \infty} \frac{1}{n} = 0\)

(b) However, even though \((\frac{1}{n})\) is \textit{convergent}, it is not \textit{Cauchy}: \(\lim_{n,m \to \infty} d(\frac{1}{n}, \frac{1}{m}) = 1 \neq 0\)

8. The \textit{distance function} \(d\) is \textit{discontinuous} in \((X, d)\):

\[
\lim_{n \to \infty} d\left(\frac{1}{n}, 2 - \frac{1}{n}\right) = 1 \neq 2 = d(0, 2) = d\left(\lim_{n \to \infty} \frac{1}{n}, \lim_{n \to \infty} \left(2 - \frac{1}{n}\right)\right).
\]

\textbf{Example B.3.} \(\text{28}\) The function \(d(x, y) \in \mathbb{R}^{\mathbb{R} \times \mathbb{R}}\) such that

\[
d(x, y) \triangleq \begin{cases} 
2|x - y| & \forall (x, y) \in \{(0, 1), (1, 0)\} \\
|x - y| & \text{otherwise}
\end{cases} \quad \text{(dilated Euclidean)}
\]

is a \textit{distance} on \(\mathbb{R}\). Note some characteristics of the \textit{distance space} \((\mathbb{R}, d)\):

1. \((\mathbb{R}, d)\) is \textit{not a metric space} because \(d\) does not satisfy the \textit{triangle inequality}:

\[
d(0, 1) \triangleq 2|0 - 1| = 2 \nless 1 = |0 - \frac{1}{2}| + |\frac{1}{2} - 1| \triangleq d(0, \frac{1}{2}) + d(\frac{1}{2}, 1)
\]

2. The function \(d\) is \textit{discontinuous}:

\[
\lim_{\to \infty} d(1 - \frac{1}{n}, \frac{1}{n}) \triangleq \lim_{\to \infty} \{1 - \frac{1}{n} - \frac{1}{n}\} = 1 \neq 2 = 2|0 - 1| \triangleq d(0, 1) = d\left(\lim_{n \to \infty} \{1 - \frac{1}{n}\}, \lim_{n \to \infty} \{\frac{1}{n}\}\right).
\]

3. In \((X, d)\), \textit{open balls} are \textit{open}:

(a) \(p(x, y) \triangleq |x - y|\) is a \textit{metric} and thus all open balls in that do not contain both 0 and 1 are \textit{open}.

(b) By Example \(D.3\) (page 157), \(q(x, y) \triangleq 2|x - y|\) is also a \textit{metric} and thus all open balls containing 0 and 1 only are \textit{open}.

(c) The only question remaining is with regards to open balls that contain 0, 1 and some other element(s) in \(\mathbb{R}\). But even in this case, open balls are still open. For example:

\[
B(-1, 2) = (-1 : 2) = (-1 : 1) \cup (1 : 2)
\]

Note that both \((-1 : 1)\) and \((1 : 2)\) are \textit{open}, and thus by Theorem \(B.2\) (page 126), \(B(-1, 2)\) is \textit{open} as well.

4. By item \((3)\) and Theorem \(B.3\) (page 128), the \textit{open balls} of \((\mathbb{R}, d)\) do form a \textit{base} for a \textit{topology} on \(\mathbb{R}\).

---

\(\text{28}\) The distance function \(d\) and item \((2)\) page 133 can in essence be found in \(\equiv\) Blumenthal (1953) page 9.

Daniel J. Greenhoe
5. In \((X, d)\), the limits of convergent sequences are unique. This is demonstrated in Example C.3 (page 144) using additional structure developed in Appendix C.

6. In \((X, d)\), convergent sequences are Cauchy. This is also demonstrated in Example C.3 (page 144).

The distance functions in Example B.1 (page 131)—Example B.3 (page 133) were all discontinuous. In the absence of the triangle inequality and in light of these examples, one might try replacing the triangle inequality with the weaker requirement of continuity. However, as demonstrated by the next example, this also leads to an arguably disastrous result.

Example B.4. 29 The function \(d \in \mathbb{R}^{\mathbb{R}^2}\) such that \(d(x, y) \triangleq (x - y)^2\) is a distance on \(\mathbb{R}\).

Note some characteristics of the distance space \((\mathbb{R}, d)\):

1. \((\mathbb{R}, d)\) is not a metric space because the triangle inequality does not hold:
   \[
d(0, 2) \triangleq (0 - 2)^2 = 4 \not< 2 = (0 - 1)^2 + (1 - 2)^2 \triangleq d(0, 1) + d(1, 2)
   \]

2. The distance function \(d\) is continuous in \((X, d)\). This is demonstrated in the more general setting of Appendix C in Example C.4 (page 145).

3. Calculating the length of curves in \((X, d)\) leads to a paradox:30
   
   (a) Partition \([0 : 1]\) into \(2^N\) consecutive line segments connected at the points 
   \[
   \left\{0, \frac{1}{2^N}, \frac{2}{2^N}, \frac{3}{2^N}, \ldots, \frac{2^N-1}{2^N}, 1\right\}
   \]
   (b) Then the distance, as measured by \(d\), between any two consecutive points is
   \[
d(p_n, p_{n+1}) \triangleq (p_n - p_{n+1})^2 = \left(\frac{1}{2^N}\right)^2 = \frac{1}{2^{2N}}
   \]
   (c) But this leads to the paradox that the total length of \([0 : 1]\) is 0:
   \[
   \lim_{N \to \infty} \sum_{n=0}^{2^N-1} \frac{1}{2^{2N}} = \lim_{N \to \infty} \frac{2^N}{2^{2N}} = \lim_{N \to \infty} \frac{1}{2^N} = 0
   \]

---

30 This is the method of “inscribed polygons” for calculating the length of a curve and goes back to Archimedes: Brunschwig et al. (2003) page 26, Walmsley (1920), page 200 (§158),
APPENDIX C

POWER DISTANCE SPACES

C.1 Definitions

This paper introduces a new relation called the power triangle inequality (Definition C.2 page 135). It is a generalization of other common relations, including the triangle inequality (Definition C.3 page 136). The power triangle inequality is defined in terms of a function herein called the power triangle function (next definition). This function is a special case of the power mean with $N = 2$ and $\lambda_1 = \lambda_2 = \frac{1}{2}$ (Definition D.15 page 164). Power means have the attractive properties of being continuous and strictly monotone with respect to a free parameter $p \in \mathbb{R}^+$ (Theorem D.14 page 164). This fact is inherited and exploited by the power triangle inequality (Corollary C.1 page 136).

**Definition C.1.** Let $(X, d)$ be a distance space (Definition B.1 page 125). Let $\mathbb{R}^+$ be the set of all positive real numbers and $\mathbb{R}^*$ be the set of extended real numbers (Definition 1.2 page 5). The power triangle function $\tau$ on $(X, d)$ is defined as

$$
\tau(p, \sigma; x, y, z; d) \triangleq 2\sigma \left( \frac{1}{2}d^p(x, z) + \frac{1}{2}d^p(z, y) \right)^\frac{1}{p} \quad \forall (p, \sigma) \in \mathbb{R}^* \times \mathbb{R}^+. \quad x, y, z \in X
$$

**Remark C.1.** In the field of probabilistic metric spaces, a function called he triangle function was introduced by Sherstnev in 1962. However, the power triangle function as defined in this present paper is not a special case of (is not compatible with) the triangle function of Sherstnev. Another definition of triangle function has been offered by Bessenyei in 2014 with special cases of $\Phi(u, v) \triangleq c(u + v)$ and $\Phi(u, v) \triangleq (u^p + v^p)^\frac{1}{p}$, which are similar to the definition of power triangle function offered in this present paper.

**Definition C.2.** Let $(X, d)$ be a distance space. Let $2^{XXX}$ be the set of all trinomial relations (Definition 1.5 page 6) on $X$. A relation $\mathcal{O}(p, \sigma; d)$ in $2^{XXX}$ is a power triangle inequality on $(X, d)$ if

$$
\mathcal{O}(p, \sigma; d) \triangleq \left\{ (x, y, z) \in X^3 \mid d(x, y) \leq \tau(p, \sigma; x, y, z; d) \right\} \quad \text{for some } (p, \sigma) \in \mathbb{R}^* \times \mathbb{R}^+.
$$

The tuple $(X, d, p, \sigma)$ is a power distance space and $d$ a power distance or power distance function if $(X, d)$ is a distance space in which the triangle relation $\mathcal{O}(p, \sigma; d)$ holds.

The power triangle function can be used to define some standard inequalities (next definition). See Corollary C.2 (page 136) for some justification of the definitions.

---

1 Sherstnev (1962), page 4, Schweizer and Sklar (1983) page 9 ((1.6.1)–(1.6.4)), Bessenyei and Pales (2014) page 2
Definition C.3.  \(^2\) Let \(\varnothing(p, \sigma; d)\) be a power triangle inequality on a distance space \((X, d)\).

1. \(\varnothing(\infty, \frac{2}{3}; d)\) is the \(\sigma\)-inframetric inequality
2. \(\varnothing(\infty, \frac{1}{2}; d)\) is the inframetric inequality
3. \(\varnothing(\ 2, \frac{2}{3}; d)\) is the quadratic inequality
4. \(\varnothing(\ 1, \sigma; d)\) is the relaxed triangle inequality
5. \(\varnothing(\ 1, \ 1; d)\) is the triangle inequality
6. \(\varnothing(\ \frac{1}{2}, \ 2; d)\) is the square mean root inequality
7. \(\varnothing(\ 0, \frac{1}{2}; d)\) is the geometric inequality
8. \(\varnothing(\ -1, \frac{2}{3}; d)\) is the harmonic inequality
9. \(\varnothing(\ -\infty, \frac{1}{2}; d)\) is the minimal inequality

Definition C.4.  \(^3\) Let \((X, d)\) be a distance space (Definition B.1 page 125).

1. \((X, d)\) is a metric space if the triangle inequality holds in \(X\).
2. \((X, d)\) is a near metric space if the relaxed triangle inequality holds in \(X\).
3. \((X, d)\) is an inframetric space if the inframetric inequality holds in \(X\).
4. \((X, d)\) is a \(\sigma\)-inframetric space if the \(\sigma\)-inframetric inequality holds in \(X\).

C.2 Properties

C.2.1 Relationships of the power triangle function

Corollary C.1. Let \(\tau(p, \sigma; x, y, z; d)\) be the power triangle function (Definition C.1 page 135) in the distance space (Definition B.1 page 125) \((X, d)\). Let \((\mathbb{R}, |·|, \leq)\) be the ordered metric space with the usual ordering relation \(\leq\) and usual metric \(|·|\) on \(\mathbb{R}\). The function \(\tau(p, \sigma; x, y, z; d)\) is continuous and strictly monotone in \((\mathbb{R}, |·|, \leq)\) with respect to both the variables \(p\) and \(\sigma\).

Proof:

1. Proof that \(\tau(p, \sigma; x, y, z; d)\) is continuous and strictly monotone with respect to \(p\): This follows directly from Theorem D.14 (page 164).

2. Proof that \(\tau(p, \sigma; x, y, z; d)\) is continuous and strictly monotone with respect to \(\sigma\):

\[
\tau(p, \sigma; x, y, z; d) \triangleq 2\sigma \frac{\left[ \frac{1}{2} d^p(x, z) + \frac{1}{2} d^p(z, y) \right]}{f(p, x, y, z)}
\]

by definition of \(\tau\) (Definition C.1 page 135)

where \(f\) is defined as above

\[\implies \tau\ is\ affine\ with\ respect\ to\ \sigma\]

\[\implies \tau\ is\ continuous\ and\ strictly\ monotone\ with\ respect\ to\ \sigma:\]

Corollary C.2. Let \(\tau(p, \sigma; x, y, z; d)\) be the power triangle function in the distance space (Definition B.1 page 125) \((X, d)\).

</Ref>

</Ref>
Figure C.1: \( \sigma = \frac{1}{2}(2^{\frac{1}{p}}) = 2^{-1} \) or \( p = \frac{\ln 2}{\ln(2\sigma)} \) (see Lemma C.1 page 137, Lemma C.2 page 139, Corollary C.6 page 140, Corollary C.5 page 139, and Theorem C.5 page 142).

\[
\tau(p, \sigma; x, y, z; d) = \begin{cases} 
2\sigma \max \{d(x, z), d(y, z)\} & \text{for } p = \infty, \\
2\sigma \left[ \frac{1}{d(x, z)} + \frac{1}{d(y, z)} \right] & \text{for } p = 0, \\
2\sigma \sqrt{d(x, z) + d(y, z)} & \text{for } p = 1, \\
4\sigma \left[\frac{1}{d(x, z)} + \frac{1}{d(y, z)} \right]^{-1} & \text{for } p = -1, \\
2\sigma \min \{d(x, z), d(y, z)\} & \text{for } p = -\infty, 
\end{cases}
\]

\( \rho \): PROOF: These follow directly from Theorem D.14 (page 164).

\( \bowtie \)

Corollary C.3. \emph{Let} \((X, d)\) \emph{be a distance space}.

\[
2\sigma \min \{d(x, z), d(z, y)\} \leq 4\sigma \left[\frac{1}{d(x, z)} + \frac{1}{d(z, y)} \right]^{-1} \leq 2\sigma \sqrt{d(x, z) + d(z, y)} \leq 2\sigma \max \{d(x, z), d(z, y)\}
\]

\( \rho \): PROOF: These follow directly from Corollary D.2 (page 167).

\( \bowtie \)

C.2.2 Properties of power distance spaces

The \emph{power triangle inequality} property of a power distance space axiomatically endows a metric with an upper bound. Lemma C.1 (next) demonstrates that there is a complementary lower bound somewhat similar in form to the \emph{power triangle inequality} upper bound. In the special case where \( 2\sigma = 2^{\frac{1}{p}} \), the lower bound helps provide a simple proof of the \emph{continuity} of a large class of \emph{power distance functions} (Theorem C.5 page 142). The inequality \( 2\sigma \leq 2^{\frac{1}{p}} \) is a simple relation in this text and appears repeatedly in this appendix; it appears as an inequality in Lemma C.2 (page 139), Corollary C.5 (page 139) and Corollary C.6 (page 140), and as an equality in Lemma C.1 (next) and Theorem C.5 (page 142). It is plotted in Figure C.1 (page 137).

Lemma C.1. \footnote{in \emph{metric space} \((p, \sigma) = (1, 1)\): \( \approx \) Dieudonné (1969) page 28, \( \approx \) Michel and Herget (1993) page 266, \( \approx \) Berberian (1961) page 37 (Theorem II.4.1)} \emph{Let} \((X, d, p, \sigma)\) \emph{be a POWER TRIANGLE TRIANGLE SPACE} (Definition C.2 page 135). \emph{Let} \(|\cdot|\) \emph{be the absolute value function} (Definition 1.28 page 11). \emph{Let} \(\max \{x, y\}\) \emph{be the maximum and} \(\min \{x, y\}\) \emph{the minimum of any} \(x, y \in \mathbb{R}^+\). \emph{Then}, for all \((p, \sigma) \in \mathbb{R}^+ \times \mathbb{R}^+\),

\begin{enumerate}
\item \( d^p(x, y) \geq \max \left\{ 0, \frac{2}{(2\sigma p)} d^p(x, z) - d^p(y, z), \frac{2}{(2\sigma p)} d^p(y, z) - d^p(z, x) \right\} \quad \forall x, y, z \in X \) \quad \text{and}
\item \( d(x, y) \geq |d(x, z) - d(z, y)| \quad \text{if} \ p \neq 0 \quad \text{and} \quad 2\sigma = 2^{\frac{1}{p}} \quad \forall x, y, z \in X. \)
\end{enumerate}
PROOF:

1. lemma: \( \frac{2}{(2\sigma)^p} d^p(x, z) - d^p(y, z) \leq d^p(x, y) \forall (p, \sigma) \in \mathbb{R}^* \times \mathbb{R}^+ \): Proof:

\[
\frac{2}{(2\sigma)^p} d^p(x, z) - d^p(y, z) \leq \frac{2}{(2\sigma)^p} \left[ \frac{1}{p} d^p(x, y) + \frac{1}{p} d^p(y, z) \right]^{\frac{1}{p}} - d^p(y, z) \quad \text{by power triangle inequality}
\]
\[
= \frac{2}{(2\sigma)^p} \left[ \frac{1}{p} d^p(x, y) + \frac{1}{p} d^p(y, z) \right] - d^p(y, z)
\]
\[
= \left[ d^p(x, y) + d^p(y, z) \right] - d^p(y, z)
\]
\[
= d^p(x, y)
\]

2. Proof for \((p, \sigma) \in \mathbb{R}^* \times \mathbb{R}^+\) case:

\[
d^p(x, y) \geq \frac{2}{(2\sigma)^p} d^p(x, z) - d^p(y, z) \quad \text{by (1) lemma}
\]
\[
d^p(x, y) = d^p(y, x) \geq \frac{2}{(2\sigma)^p} d^p(y, z) - d^p(x, z) \quad \text{by commutative property of } d \text{ and (1) lemma}
\]
\[
d^p(x, y) \geq 0 \quad \text{by non-negative property of } d \quad \text{(Definition B.1 page 125)}
\]

The rest follows because \(g(x) \triangleq x^{\frac{1}{p}}\) is strictly monotone in \(\mathbb{R}^+\).

3. Proof for \(2\sigma = 2^{\frac{1}{p}}\) case:

\[
d(x, y) \geq \max \left\{ 0, \frac{2}{(2\sigma)^p} d^p(x, z) - d^p(y, z), \frac{2}{(2\sigma)^p} d^p(y, z) - d^p(x, z) \right\}^{\frac{1}{p}} \quad \text{by item (2) (page 138)}
\]
\[
= \max \{ 0, d(x, z) - d(z, y), d(y, z) - d(z, x) \} \quad \text{by } 2\sigma = 2^{\frac{1}{p}} \text{ hypothesis } \iff \frac{2}{(2\sigma)^p} = 1
\]
\[
= \max \{ 0, (d(x, z) - d(z, y)), -(d(x, z) - d(z, y)) \} \quad \text{by symmetric property of } d
\]
\[
= |(d(x, z) - d(z, y))|
\]

\[\implies \]

Theorem C.1. Let \((X, d, p, \sigma)\) be a POWER DISTANCE SPACE (Definition C.2 page 135). Let \(B\) be an OPEN BALL (Definition B.4 page 126) on \((X, d)\). Then for all \((p, \sigma) \in (\mathbb{R}^* \setminus \{0\}) \times \mathbb{R}^+\),

\[
\left\{ \begin{array}{ll}
A & \quad 2\sigma \leq 2^{\frac{1}{p}} \quad \text{and} \\
B & \quad q \in B(\theta, r)
\end{array} \right\} \quad \implies \quad \left\{ \begin{array}{ll}
1 & \quad \exists r_q \in \mathbb{R}^+ \quad \text{such that} \\
2 & \quad B(q, r_q) \subseteq B(\theta, r)
\end{array} \right\} \quad \implies \quad \left\{ \begin{array}{ll}
B & \quad q \in B(\theta, r)
\end{array} \right\}
\]

\[\implies \]

1. lemma:

\[
q \in B(\theta, r) \iff d(\theta, q) < r \quad \text{by definition of open ball (Definition B.4 page 126)}
\]
\[
\iff 0 < r - d(\theta, q) \quad \text{by field property of real numbers}
\]
\[
\iff \exists r_q \in \mathbb{R}^+ \quad \text{such that} \quad 0 < r_q < r - d(\theta, q) \quad \text{by The Archimedean Property}^5
\]

---

^5 Aliprantis and Burkinshaw (1998) page 17 (Theorem 3.3 ("The Archimedean Property") and Theorem 3.4), Zorich (2004) page 53 (6 ("The principle of Archimedes") and 7)
2. Proof that \((A), (B) \implies (1)\):

\[
B(q, r_q) \triangleq \{ x \in X | d(q, x) < r_q \}
\]

by definition of open ball (Definition B.4 page 126)

\[
= \{ x \in X | d^p(q, x) < r_q^p \in \mathbb{R}^+ \}
\]

because \(f(x) \triangleq x^p\) is monotone

\[
\subseteq \{ x \in X | d^p(q, x) < r^p - d^p(\theta, q) \}
\]

by hypothesis B and (1) lemma page 138

\[
= \{ x \in X | d^p(\theta, q) + d^p(q, x) < r^p \}
\]

by field property of real numbers

\[
= \{ x \in X | d^p(\theta, q) + d^p(q, x) \frac{1}{2} < r \}
\]

by hypothesis A which implies \(2^{1-p} \sigma \leq 1\)

\[
\triangleq \{ x \in X | \tau(p, \sigma, \theta, x, q) < r \}
\]

by definition of \(\tau\) (Definition C.1 page 135)

\[
\subseteq \{ x \in X | d(\theta, x) < r \}
\]

by definition of \((X, d, p, \sigma)\) (Definition C.2 page 135)

\[
\triangleq B(\theta, r)
\]

by definition of open ball (Definition B.4 page 126)

3. Proof that \((B) \iff (1)\):

\[
q \in \{ x \in X | d(q, x) = 0 \}
\]

by nondegenerate property (Definition B.1 page 125)

\[
\subseteq \{ x \in X | d(q, x) < r_q \}
\]

because \(r_q > 0\)

\[
\triangleq B(q, r_q)
\]

by definition of open ball (Definition B.4 page 126)

\[
\subseteq B(\theta, r)
\]

by hypothesis 2

\[\blacksquare\]

**Corollary C.4.** Let \((X, d, p, \sigma)\) be a power distance space. Then for all \((p, \sigma) \in (\mathbb{R}^+ \setminus \{0\}) \times \mathbb{R}^+\),

\[
\left\{ 2\sigma \leq 2^{\frac{1}{p}} \right\} \implies \text{\{ every open ball in } (X, d) \text{ is open\}}
\]

**Proof:** This follows from Theorem C.1 (page 138) and Theorem B.3 (page 128).

\[\blacksquare\]

**Corollary C.5.** Let \((X, d, p, \sigma)\) be a power distance space. Let \(B\) be the set of all open balls in \((X, d)\). Then for all \((p, \sigma) \in (\mathbb{R}^+ \setminus \{0\}) \times \mathbb{R}^+\),

\[
\left\{ 2\sigma \leq 2^{\frac{1}{p}} \right\} \implies \text{\{ \(B\) is a base for } (X, T) \text{ \}}
\]

**Proof:**

1. The set of all open balls in \((X, d)\) is a base for \((X, T)\) by Corollary C.4 (page 139) and Theorem D.16 (page 168).

2. \(T\) is a topology on \(X\) by Definition D.17 (page 168).

\[\blacksquare\]

Lemma C.2 (next) demonstrates that every point in an open set is contained in an open ball that is contained in the original open set (see also Figure C.2 page 140).

**Lemma C.2.** Let \((X, d, p, \sigma)\) be a power distance space. Let \(B\) be an open ball on \((X, d)\). Then for all \((p, \sigma) \in (\mathbb{R}^+ \setminus \{0\}) \times \mathbb{R}^+\),

\[
\left\{ A. \ \ 2\sigma \leq 2^{\frac{1}{p}} \text{ and } B. \ U\text{ is open in } (X, d) \right\} \implies \left\{ i. \ \forall x \in U, \ \exists r \in \mathbb{R}^+ \text{ such that } B(x, r) \subseteq U \right\} \implies \left\{ B. \ U \text{ is open in } (X, d) \right\}
\]

**Proof:**
1. Proof that for \( (A), (B) \implies (1) \):

\[
U = \bigcup \{ B(x, r) \mid B(x, r) \subseteq U \} \quad \text{by left hypothesis and Corollary C.5 page 139}
\]

because \( x \) must be in one of those balls in \( U \)

2. Proof that \((B) \iff (1)\) case:

\[
U = \bigcup \{ x \in X \mid x \in U \} = \bigcup \{ B(x, r) \mid x \in U \text{ and } B(x, r) \subseteq U \} \quad \text{by hypothesis (1)}
\]

\[\implies U \text{ is open} \quad \text{by Corollary C.5 page 139 and Corollary B.1 page 127}\]

**Corollary C.6.** Let \((X, d, p, \sigma)\) be a power distance space. Let \(B\) be an open ball on \((X, d)\). Then for all \((p, \sigma) \in (\mathbb{R} \setminus \{0\}) \times \mathbb{R}^+\),

\[
\left\{ 2\sigma \leq 2^{\frac{1}{2}} \right\} \implies \left\{ \text{every open ball } B(x, r) \text{ in } (X, d) \text{ is open} \right\}
\]

**Proof:**

The union of any set of open balls is open by Corollary C.5 page 139

\[\implies \text{the union of a set of just one open ball is open} \]

\[\implies \text{every open ball is open.}\]

**Theorem C.2.** Let \((X, d, p, \sigma)\) be a power distance space. Let \((X, T)\) be a topological space induced by \((X, d)\). Let \(\{x_n \in X\}_{n \in \mathbb{Z}}\) be a sequence in \((X, d)\).

\[
\{x_n\} \text{ converges to a limit } x \quad \iff \quad \left\{ \text{for any } \varepsilon \in \mathbb{R}^+, \text{ there exists } N \in \mathbb{Z} \text{ such that for all } n > N, \quad d(x_n, x) < \varepsilon \right\}
\]

**Proof:**

\[
\{x_n\} \to x \iff x_n \in U \quad \forall U \in N_x, \ n > N \quad \text{by Definition D.20 page 170}
\]

\[\iff \exists B(x, \varepsilon) \text{ such that } x_n \in B(x, \varepsilon) \forall n > N \quad \text{by Lemma C.2 page 139}\]

\[\iff d(x_n, x) < \varepsilon \quad \text{by Definition B.4 page 126}\]

---

6 in metric space ((p, \sigma) = (1, 1)): \(\equiv\) Rosenlicht (1968) pages 40–41, \(\equiv\) Aliprantis and Burkinshaw (1998) page 35

7 in metric space: \(\equiv\) Rosenlicht (1968) page 45, \(\equiv\) Giles (1987) page 37 (3.2 Definition)
In distance spaces (Definition B.1 page 125), not all convergent sequences are Cauchy (Example B.2 page 132). However in a distance space with any power triangle inequality (Definition C.2 page 135), all convergent sequences are Cauchy (next theorem).

**Theorem C.3.** Let \((X, d, p, \sigma)\) be a power distance space. Let \(B\) be an open ball on \((X, d)\). For any \((p, \sigma) \in \mathbb{R}^+ \times \mathbb{R}^+\),

\[
\begin{align*}
\{ (x_n) \text{ is convergent in } (X, d) \} & \quad \Rightarrow \quad \{ (x_n) \text{ is Cauchy in } (X, d) \} \\
& \quad \Rightarrow \quad \{ (x_n) \text{ is bounded in } (X, d) \}
\end{align*}
\]

**Proof:**

1. Proof that convergent \(\Rightarrow\) Cauchy:

\[
d(x_n, x_m) \leq \tau(p, \sigma; x_n, x_m, x) \quad \text{by definition of power triangle inequality (Definition C.2 page 135)}
\]

\[
\triangleq 2\sigma \left[ \frac{1}{2} d^p(x_n, x) + \frac{1}{2} d^p(x_m, x) \right]^\frac{1}{p}
\]

\[
< 2\sigma \left[ \frac{1}{2} \epsilon^p + \frac{1}{2} \epsilon^p \right]^\frac{1}{p}
\]

\[
= 2\sigma \epsilon
\]

\[
\Rightarrow \text{ Cauchy}
\]

\[
d(x_n, x_m) \leq \tau(\infty, \sigma; x_n, x_m, x) \quad \text{by definition of power triangle inequality at } p = \infty
\]

\[
= 2\sigma \max \{ d(x_n, x), d(x_m, x) \}
\]

\[
= 2\sigma \max \{ \epsilon, \epsilon \}
\]

\[
= 2\sigma \epsilon
\]

\[
\text{by Corollary C.2 (page 136)}
\]

\[
\text{by convergent hypothesis (Definition D.20 page 170)}
\]

\[
\text{by definition of max}
\]

\[
d(x_n, x_m) \leq \tau(-\infty, \sigma; x_n, x_m, x) \quad \text{by definition of power triangle inequality at } p = -\infty
\]

\[
= 2\sigma \min \{ d(x_n, x), d(x_m, x) \}
\]

\[
= 2\sigma \min \{ \epsilon, \epsilon \}
\]

\[
= 2\sigma \epsilon
\]

\[
\text{by Corollary C.2 (page 136)}
\]

\[
\text{by convergent hypothesis (Definition D.20 page 170)}
\]

\[
\text{by definition of min}
\]

\[
d(x_n, x_m) \leq \tau(0, \sigma; x_n, x_m, x) \quad \text{by definition of power triangle inequality at } p = 0
\]

\[
= 2\sigma \sqrt{d(x_n, x) / d(x_m, x)}
\]

\[
= 2\sigma \sqrt{\epsilon / \epsilon}
\]

\[
= 2\sigma \epsilon
\]

\[
\text{by Corollary C.2 (page 136)}
\]

\[
\text{by convergent hypothesis (Definition D.20 page 170)}
\]

\[
\text{by property of } \mathbb{R}
\]

2. Proof that Cauchy \(\Rightarrow\) bounded: by Proposition B.1 (page 129).

**Theorem C.4.** Let \((X, d, p, \sigma)\) be a power distance space. Let \(f \in \mathbb{Z}^\mathbb{Z}\) be a strictly monotone function such that \(f(n) < f(n + 1)\). For any \((p, \sigma) \in \mathbb{R}^+ \times \mathbb{R}^+
\]

\[
\begin{align*}
\{ (x_n)_{n \in \mathbb{Z}} \text{ is Cauchy and } \} & \quad \Rightarrow \quad \{ (x_n)_{n \in \mathbb{Z}} \text{ is convergent.} \}
\end{align*}
\]

\[\text{in metric space: } \text{Giles (1987) page 49 (Theorem 3.30), Rosenlicht (1968) page 51, Apostol (1975) pages 72–73 (Theorem 4.6)}\]

\[\text{in metric space: } \text{Rosenlicht (1968) page 52}\]
PROOF:

\[ d(x_n, x) = d(x, x_n) \]
\[ \leq \tau(p, \sigma; x, x_n, x_{t(n)}) \]
\[ \triangleq 2\sigma \left[ \frac{1}{2} d^p(x, x_{t(n)}) + \frac{1}{2} d^p(x_{t(n)}, x_n) \right]^\frac{1}{p} \]
\[ = 2\sigma \left[ \frac{1}{2} \epsilon_p + \frac{1}{2} d^p(x_{t(n)}, x_n) \right]^\frac{1}{p} \]
\[ = 2\sigma \epsilon \]
\[ \implies \text{convergent} \]

by symmetric property of \( d \)

by definition of power triangle inequality (Definition C.2 page 135)

by definition of power triangle function (Definition C.1 page 135)

by left hypothesis 2

by left hypothesis 1

by definition of convergent (Definition D.20 page 170)

Theorem C.5. \( ^{10} \) Let \( (X, d, p, \sigma) \) be a power distance space. Let \( (\mathbb{R}, q) \) be a metric space of real numbers with the usual metric \( q(x, y) \triangleq |x - y| \). Then

\[ \{ 2\sigma = 2^\frac{1}{p} \} \implies \{ \text{d is continuous in } (\mathbb{R}, q) \} \]

PROOF:

\[ |d(x, y) - d(x_n, y_n)| \leq |d(x, y) - d(x_n, y)| + |d(x_n, y) - d(x_n, y_n)| \]
\[ = |d(x, y) - d(y, x_n)| + |d(y, x_n) - d(x_n, y_n)| \]
\[ \leq d(x, x_n) + d(y, x_n) \]
\[ = 0 \quad \text{as } n \to \infty \]

by triangle inequality of \((\mathbb{R}, |x - y|)\)

by commutative property of \( d \) (Definition B.1 page 125)

by \( 2\sigma = 2^\frac{1}{p} \) and Lemma C.1 (page 137)

In distance spaces and topological spaces, limits of convergent sequences are in general not unique (Example B.1 page 131, Example D.16 page 170). However Theorem C.6 (next) demonstrates that, in a power distance space, limits are unique.

Theorem C.6 (Uniqueness of limit). \( ^{11} \) Let \( (X, d, p, \sigma) \) be a power distance space. Let \( x, y, \in X \) and let \( (x_n, \in X) \) be an \( X \)-valued sequence.

\[ \{ \begin{array}{c}
1. \quad \{ (x_n, (y_n)) \to (x, y) \} \\
2. \quad (p, \sigma) \in \mathbb{R}^+ \times \mathbb{R}^+ \end{array} \} \implies \{ x = y \} \]

PROOF:

1. lemma: Proof that for all \((p, \sigma) \in \mathbb{R}^+ \times \mathbb{R}^+ \) and for any \( \epsilon \in \mathbb{R}^+ \), there exists \( N \) such that \( d(x, y) < 2\sigma \epsilon \):

\[ d(x, y) \leq \tau(p, \sigma; x, y, x_n) \]
\[ \triangleq 2\sigma \left[ \frac{1}{2} d^p(x, x_n) + \frac{1}{2} d^p(x_n, y) \right]^\frac{1}{p} \]
\[ < 2\sigma \left[ \frac{1}{2} \epsilon_p + \frac{1}{2} d^p(x_n, y) \right]^\frac{1}{p} \]
\[ = 2\sigma \epsilon \]
\[ d(x, y) \leq \tau(\infty, \sigma; x, y, x_n) \]
\[ = 2\sigma \max \{ d(x, x_n), d(x_n, y) \} \]

by definition of power triangle inequality (Definition C.2 page 135)

by definition of power triangle function (Definition C.1 page 135)

by left hypothesis and for \( p \in \mathbb{R}^+ \setminus [-\infty, 0, \infty] \)

by definition of power triangle inequality at \( p = \infty \)

by Corollary C.2 (page 136)

\(^{10}\) in metric space \((p, \sigma) = (1, 1) \) case: \( ^{\triangleright} \) Berberian (1961) page 37 (Theorem II.4.1)

\(^{11}\) in metric space: \( ^{\triangleright} \) Rosenlicht (1968) page 46, \( ^{\triangleright} \) Thomson et al. (2008) page 32 (Theorem 2.8)
\[ d(x, y) \leq \tau(-\infty, \sigma; x, y, x_n) = 2\sigma \min \{ d(x, x_n), d(x_n, y) \}, \]
\[ < 2\sigma \epsilon \] by left hypothesis

by definition of power triangle inequality at \( p = -\infty \)

by Corollary C.2 (page 136)

by left hypothesis

by Corollary C.2 (page 136)

by left hypothesis

by property of real numbers

2. Proof that \( x = y \) (proof by contradiction):

\[ x \neq y \implies d(x, y) \neq 0 \]
\[ \implies d(x, y) > 0 \]
\[ \implies \exists \epsilon \text{ such that } d(x, y) > 2\epsilon \]
\[ \implies \text{contradiction to (1) lemma page 142} \]
\[ \implies d(x, y) = 0 \]
\[ \implies x = y \]

\[ \therefore \]

C.3 Examples

It is not always possible to find a triangle relation (Definition C.2 page 135) \( \odot(p, \sigma; d) \) that holds in every distance space (Definition B.1 page 125), as demonstrated by Example C.1 and Example C.2 (next two examples).

Example C.1. Let \( d(x, y) \in \mathbb{R}^{x \times R} \) be defined such that

\[
 d(x, y) \triangleq \begin{cases} 
 y & \forall (x, y) \in \{4\} \times \{0 : 2\} \quad \text{(vertical half-open interval)} \\
 x & \forall (x, y) \in \{0 : 2\} \times \{4\} \quad \text{(horizontal half-open interval)} \\
 |x - y| & \text{otherwise} \quad \text{(Euclidean)}
\end{cases}
\]

Note the following about the pair \((\mathbb{R}, d)\):

1. By Example B.1 (page 131), \((\mathbb{R}, d)\) is a distance space, but not a metric space—that is, the triangle relation \( \odot(1, 1; d) \) does not hold in \((\mathbb{R}, d)\).

2. Observe further that \((\mathbb{R}, d)\) is not a power distance space. In particular, the triangle relation \( \odot(p, \sigma; d) \) does not hold in \((\mathbb{R}, d)\) for any finite value of \( \sigma \) (does not hold for any \( \sigma \in \mathbb{R}^+ \)):

\[
 d(0, 4) = 4 \not\leq 0 = \lim_{\epsilon \rightarrow 0} 2\sigma \epsilon = \lim_{\epsilon \rightarrow 0} 2\sigma [\frac{1}{2}|0 - \epsilon|_p + \frac{1}{2}\epsilon^p]^{\frac{1}{p}} \\
 \leq \lim_{\epsilon \rightarrow 0} 2\sigma [\frac{1}{2}d^p(0, \epsilon) + \frac{1}{2}d^p(\epsilon, 4)]^{\frac{1}{p}} \not\leq \lim_{\epsilon \rightarrow 0} \odot(p, \sigma; 0, 4, \epsilon; d)
\]

Example C.2. Let \( d(x, y) \in \mathbb{R}^{x \times R} \) be defined such that

\[
 d(x, y) \triangleq \begin{cases} 
 |x - y| & \text{for } x = 0 \text{ or } y = 0 \text{ or } x = y \quad \text{(Euclidean)} \\
 1 & \text{otherwise} \quad \text{(discrete)}
\end{cases}
\]

Note the following about the pair \((\mathbb{R}, d)\):

1. By Example B.2 (page 132), \((\mathbb{R}, d)\) is a distance space, but not a metric space—that is, the triangle relation \( \odot(1, 1; d) \) does not hold in \((\mathbb{R}, d)\).
2. Observe further that \((\mathbb{R}, d)\) is not a power distance space—that is, the triangle relation \(\odot(p, \sigma; d)\) does not hold in \((\mathbb{R}, d)\) for any value of \((p, \sigma) \in \mathbb{R}^\ast \times \mathbb{R}^+\).

(a) Proof that \(\odot(p, \sigma; d)\) does not hold for any \((p, \sigma) \in \{\infty\} \times \mathbb{R}^+\):

\[
\lim_{n, \ell \to \infty} d(\frac{\ell}{n}, \frac{\ell}{n}) \triangleq 1 \not\leq 0 = 2\sigma \max \{0, 0\} \quad \text{by definition of} \ d
\]

\[
= 2\sigma \lim_{n, \ell \to \infty} \max \{d(\frac{\ell}{n}, 0), d(0, \frac{\ell}{n})\} \quad \text{by Corollary C.2 (page 136)}
\]

\[
\geq 2\sigma \left[\frac{1}{2}d^p(\frac{\ell}{n}, 0) + \frac{1}{2}d^p(0, \frac{\ell}{n})\right]^\frac{1}{p} \quad \text{by Corollary C.1 (page 136)}
\]

\[
\triangleq \lim_{n, \ell \to \infty} \tau(p, \sigma, \frac{\ell}{n}, \frac{\ell}{n}) \quad \text{by definition of} \ \tau \ (\text{Definition C.1 page 135)}
\]

(b) Proof that \(\odot(p, \sigma; d)\) does not hold for any \((p, \sigma) \in \mathbb{R}^\ast \times \mathbb{R}^+\): By Corollary C.1 (page 136), the triangle function (Definition C.1 page 135) \(\tau(p, \sigma; x, y, z; d)\) is continuous and strictly monotone in \((\mathbb{R}, |\cdot|, \leq)\) with respect to the variable \(p\). Item 2a demonstrates that \(\odot(p, \sigma; d)\) fails to hold at the best case of \(p = \infty\), and so by Corollary C.1, it doesn't hold for any other value of \(p \in \mathbb{R}^\ast\) either.

Example C.3. Let \(d\) be a function in \(\mathbb{R}^\ast \times \mathbb{R}\) such that

\[
d(x, y) \triangleq \begin{cases} 
2|x - y| & \forall (x, y) \in \{(0, 1), (1, 0)\} \quad \text{(dilated Euclidean)} \\
|x - y| & \text{otherwise} \quad \text{(Euclidean)} 
\end{cases}
\]

Note the following about the pair \((\mathbb{R}, d)\):

1. By Example B.3 (page 133), \((\mathbb{R}, d)\) is a distance space, but not a metric space—that is, the triangle relation \(\odot(1, 1; d)\) does not hold in \((\mathbb{R}, d)\).

2. But observe further that \((\mathbb{R}, d, 1, 2)\) is a power distance space:

(a) Proof that \(\odot(1, 2; d)\) (Definition C.2 page 135) holds for all \((x, y) \in \{(0, 1), (1, 0)\}):

\[
d(0, 1) = d(0, 1) \triangleq 2|0 - 1| = 2 \quad \text{by definition of} \ d
\]

\[
\leq 2 \leq 2(0 - z) + |z - 1| \quad \forall z \in \mathbb{R} \quad \text{by definition of} \ |\cdot| \ (\text{Definition 1.28 page 11)}
\]

\[
= 2\sigma\left(\frac{1}{2}|0 - z|^p + \frac{1}{2}|z - 1|^p\right)^\frac{1}{p} \quad \forall z \in \mathbb{R} \quad \text{for} \ (p, \sigma) = (1, 2)
\]

\[
\triangleq 2\sigma\left(\frac{1}{2}d^p(0, z) + d^p(z, 1)\right)^\frac{1}{p} \quad \forall z \in \mathbb{R} \quad \text{for} \ (p, \sigma) = (1, 2) \text{ and by definition of} \ d
\]

\[
\triangleq \tau(1, 2; 0, 1, z) \quad \text{by definition of} \ \tau \ (\text{Definition C.1 page 135)}
\]

(b) Proof that \(\odot(1, 2; d)\) holds for all other \((x, y) \in \mathbb{R}^\ast \times \mathbb{R}^+\):

\[
d(x, y) \triangleq 2|x - y| \quad \text{by definition of} \ d
\]

\[
\leq (|x - z| + |z - y|) \quad \text{by property of Euclidean metric spaces}
\]

\[
= 2\sigma\left(\frac{1}{2}|0 - z|^p + \frac{1}{2}|z - 1|^p\right)^\frac{1}{p} \quad \text{for} \ (p, \sigma) = (1, 1)
\]

\[
\triangleq \tau(1, 1; x, y, z) \quad \text{by definition of} \ \tau \ (\text{Definition C.1 page 135)}
\]

\[
\leq \tau(1, 2; x, y, z) \quad \text{by Corollary C.1 (page 136)}
\]

3. In \((X, d)\), the limits of convergent sequences are unique. This follows directly from the fact that \((\mathbb{R}, d, 1, 2)\) is a power distance space (item (2) page 144) and by Theorem C.6 page 142.

4. In \((X, d)\), convergent sequences are Cauchy. This follows directly from the fact that \((\mathbb{R}, d, 1, 2)\) is a power distance space (item (2) page 144) and by Theorem C.3 page 141.
Example C.4. Let \( d \) be a function in \( \mathbb{R}^{\mathbb{R} \times \mathbb{R}} \) such that \( d(x, y) \triangleq (x - y)^2 \). Note the following about the pair \((\mathbb{R}, d)\):

1. It was demonstrated in Example B.4 (page 134) that \((\mathbb{R}, d)\) is a distance space, but that it is not a metric space because the triangle inequality does not hold.

2. However, the tuple \((\mathbb{R}, d, p, \sigma)\) is a power distance space (Definition C.2 page 135) for any \((p, \sigma) \in \mathbb{R}^+ \times [2 : \infty)\): In particular, for all \( x, y, z \in \mathbb{R} \), the power triangle inequality (Definition C.2 page 135) must hold. The “worst case” for this is when a third point \( z \) is exactly “halfway between” \( x \) and \( y \) in \( d(x, y) \); that is, when \( z = \frac{x+y}{2} \):

\[
(x - y)^2 \triangleq d(x, y) \quad \text{by definition of } d
\]

\[
\leq \tau(p, \sigma; x, y, z; d)
\]

\[
\triangleq 2\sigma \left[ \frac{1}{p}d^p(x, z) + \frac{1}{p}d^p(z, y) \right]^{\frac{1}{p}} \quad \text{by definition } \tau \text{ (Definition C.1 page 135)}
\]

\[
\triangleq 2\sigma \left[ \frac{1}{p}(x - z)^{2p} + \frac{1}{p}(z - y)^{2p} \right]^{\frac{1}{p}} \quad \text{by definition of } d
\]

\[
= 2\sigma \left[ \frac{1}{p} \left( \frac{x - y}{2} \right)^{2p} + \frac{1}{p} \left( \frac{x - y}{2} \right)^{2p} \right]^{\frac{1}{p}}
\]

\[
= 2\sigma \left[ \left( \frac{x - y}{2} \right)^{2p} + \left( \frac{x - y}{2} \right)^{2p} \right]^{\frac{1}{p}}
\]

\[
= 2\sigma \left( \left( \frac{x - y}{2} \right)^{2p} \right)^{\frac{1}{p}} = \frac{2\sigma}{4} |x - y|^2
\]

\[
\implies (p, \sigma) \in \mathbb{R}^+ \times [2 : \infty)
\]

3. The power distance function \( d \) is continuous in \((\mathbb{R}, d, p, \sigma)\) for any \((p, \sigma)\) such that \( \sigma \geq 2 \) and \( 2\sigma = p^\frac{1}{p} \). This follows directly from Theorem C.5 (page 142).
"Dirichlet alone, not I, nor Cauchy, nor Gauss knows what a completely rigorous proof is. Rather we learn it first from him. When Gauss says he has proved something it is clear; when Cauchy says it, one can wager as much pro as con; when Dirichlet says it, it is certain."

Carl Gustav Jacob Jacobi (1804–1851), Jewish-German mathematician

### D.1 Linear spaces

#### D.1.1 Structure

**Definition D.1.** Let \((\mathbb{F}, +, \cdot)\) be a FIELD. Let \(X\) be a set, let \(+\) be an OPERATOR in \(X^{\times 2}\), and let \(\otimes\) be an operator in \(X^{\times X}\). The structure \(\Omega \triangleq (X, +, \cdot, (\mathbb{F}, +, \cdot))\) is a linear space over \((\mathbb{F}, +, \cdot)\) if:

1. \(\exists 0 \in X\) such that \(x + 0 = x\) \(\forall x \in X\) (identity)
2. \(\exists 1 \in X\) such that \(x + 1 = x\) \(\forall x \in X\) (inverse)
3. \((x + y) + z = x + (y + z)\) \(\forall x, y, z \in X\) (associative)
4. \(x + y = y + x\) \(\forall x, y \in X\) (commutative)
5. \(1 \cdot x = x\) \(\forall x \in X\) (identity)
6. \(\alpha \cdot (\beta \cdot x) = (\alpha \cdot \beta) \cdot x\) \(\forall \alpha, \beta \in S\) and \(x \in X\) (associates with \(\cdot\))
7. \(\alpha \cdot (x + y) = (\alpha \cdot x) + (\alpha \cdot y)\) \(\forall \alpha \in S\) and \(x, y \in X\) (distributes over \(+\))
8. \((\alpha + \beta) \cdot x = (\alpha \cdot x) + (\beta \cdot x)\) \(\forall \alpha, \beta \in S\) and \(x \in X\) (pseudo-distributes over \(+\))

**Definition D.2.** Let \(\Omega \triangleq (X, +, \cdot, (\mathbb{F}, +, \cdot))\) be a LINEAR SPACE (Definition D.1 page 147). A set \(D \subseteq X\) is convex in \(\Omega\) if

\[
\lambda x + (1 - \lambda)y \in D \quad \forall x, y \in D \quad \forall \lambda \in (0, 1)
\]

A set is concave in \(\Omega\) if it is NOT convex in \(\Omega\).

---

1. quote: [http://lagrange.math.trinity.edu/aholder/misc/quotes.shtml](http://lagrange.math.trinity.edu/aholder/misc/quotes.shtml)
   Shaubring (2005), page 558


147
D.1.2 Metric Linear Spaces

Metric structure can be added to a linear space resulting in a metric linear space (next definition). One key difference between metric linear spaces and normed linear spaces is that the balls in a normed linear space (Definition D.4 page 149) are always convex (Definition D.2 page 147); this is not true for all metric linear spaces (Theorem D.4 page 150).

Definition D.3. ⁵ Let $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, +, \cdot), d)$. The tuple $\Omega$ is a metric linear space if
1. $d(x, y)$ is a metric in $\mathbb{F}$.
2. $d(x + z, y + z) = d(x, y)$ is a linear space and
3. $d(x + z, y + z) \leq d(x, y)$ is a translation invariant and
4. $\alpha_n \to \alpha$ and $x_n \to x$ implies $\alpha_n x_n \to \alpha x$.

Theorem D.1. ⁷ Let $\Omega \triangleq (X, +, \cdot, (\mathbb{F}, +, \cdot), d)$ be a metric linear space.

$$d(\theta, \lambda x + (1 - \lambda)y) \leq \lambda d(\theta, x) + (1 - \lambda)d(\theta, y)$$

Proof:

$$d(\theta, \lambda x + (1 - \lambda)y) \leq \lambda d(\theta, x) + (1 - \lambda)d(\theta, y) \leq \lambda r + (1 - \lambda)r$$

$$= r$$

$$\implies \lambda x + (1 - \lambda)y \in B(\theta, r)$$

$$\implies B(\theta, r) \in (X, d)$$

is convex

∀θ ∈ X.

Theorem D.2. ⁸ Let $\Omega \triangleq (X, +, \cdot, (\mathbb{R}, +, \cdot), d)$ be a real metric linear space.

$$d(x + z, y + z) = d(x, y) \forall x, y, z \in X$$

(translation invariant) and

$$\lambda d(x, y) = \lambda d(x, y) \forall x, y \in X, \lambda \in [0, 1]$$

(homogeneous)

Proof:

$$d(\theta, \lambda x + (1 - \lambda)y) = d(0, \lambda x + (1 - \lambda)y - \theta)$$

by translation invariance hypothesis

$$= d(0, \lambda(x - \theta) + (1 - \lambda)(y - \theta))$$

$$\leq d(0, \lambda(x - \theta)) + d(\lambda(x - \theta), \lambda(x - \theta) + (1 - \lambda)(y - \theta))$$

by subadditive property

$$= d(0, \lambda(x - \theta)) + d(0, 0 + (1 - \lambda)(y - \theta))$$

by translation invariance hypothesis

$$= \lambda d(0, x - \theta) + (1 - \lambda)d(0, y - \theta)$$

by homogeneous hypothesis

$$= \lambda d(0, x - \theta) + (1 - \lambda)d(\theta, y)$$

by translation invariance hypothesis

---

⁴ Bruckner et al. (1997) page 478
⁶ Some authors do not require the translation invariant property for the definition of the metric linear space, as indicated by the following references: Maddox (1989) page 90 ("Some authors...do not include translation invariance in the definition of metric linear space, since they use a theorem of Kakutani to show that a non-translation invariant metric may be replaced by a translation invariant metric which yields the same topology.") Friedman (1970) page 125 (Definition 4.1.4), Dobrowski and Mogilski (1995) page 86
⁷ Norfolk (1991), page 5
⁸ Norfolk (1991) pages 5-6, http://groups.google.com/group/sci.math/msg/a6f0a7924027957d
D.1.3 Normed Linear Spaces

**Definition D.4.** Let \((X, +, \cdot, (\mathbb{F}, +, \times))\) be a linear space (Definition D.1 page 147) and \(|\cdot| \in \mathbb{R}^F\) the absolute value function. A functional \(\|\cdot\|\) in \(\mathbb{R}^X\) is a norm if

1. \(\|x\| \geq 0\) \(\forall x \in X\) (strictly positive)
2. \(\|x\| = 0 \iff x = 0\) \(\forall x \in X\) (nondegenerate)
3. \(\|\alpha x\| = |\alpha| \|x\|\) \(\forall \alpha \in \mathbb{F}, x \in X\) (homogeneous)
4. \(\|x + y\| \leq \|x\| + \|y\|\) \(\forall x, y \in X\) (subadditive/triangle inequality).

A normed linear space is the tuple \((X, +, \cdot, (\mathbb{F}, +, \times), \|\cdot\|)\).

**Example D.1** (The usual norm). Let \(\mathbb{R}^2\) be the set of all functions with domain and range the set of real numbers \(\mathbb{R}\).

The absolute value \(|\cdot| \in \mathbb{R}^R\) is a norm.

**Example D.2** \((l_p\) norms). Let \((x_n)_{n \in \mathbb{Z}}\) be a sequence of real numbers.

\[\|\langle x_n \rangle\|_p \triangleq \left(\sum_{n \in \mathbb{Z}} |x_n|^p\right)^{\frac{1}{p}}\]

is a norm for \(p \in [1: \infty]\)

D.1.4 Relationship between metrics and norms

Metrics generated by norms

**Theorem D.3.** Let \(d \in \mathbb{R}^{X \times X}\) be a function on a real normed linear space \((X, +, \cdot, (\mathbb{F}, +, \times), \|\cdot\|)\). Let \(B(x, r) \triangleq \{y \in X | \|y - x\| < r\}\) be the open ball (Definition B.4 page 126) of radius \(r\) centered at a point \(x\).

\[d(x, y) \triangleq \|x - y\|\]

is a metric on \(X\).

The next definition defines this metric formally.

**Definition D.5.** Let \((X, +, \cdot, (\mathbb{F}, +, \times), \|\cdot\|)\) be a normed linear space (Definition D.4 page 149). The metric induced by the norm \(\|\cdot\|\) is the function \(d \in \mathbb{R}^X\) such that \(d(x, y) \triangleq \|x - y\| \forall x, y \in X\).

**Corollary D.1.** Let \(\mathcal{O} \triangleq (X, +, \cdot, (\mathbb{F}, +, \times), \|\cdot\|)\) be a normed linear space (Definition D.4 page 149). The norm \(\|\cdot\|\) is continuous in \(\mathcal{O}\).
Theorem D.4 (next) demonstrates that all open or closed balls in any normed linear space are convex. However, the converse is not true—that is, a metric not generated by a norm may still produce a ball that is convex.

Theorem D.4.¹⁴ Let \((X, +, \cdot, (\mathbb{F}, +, \times), d)\) be a metric linear space (Definition D.3 page 148). Let \(B\) be an open ball (Definition B.4 page 126),

\[
\exists \, \|\cdot\| \in \mathbb{R}^X \text{ such that } d(x, y) = \|y - x\| \quad \text{d is generated by a norm}
\]

\[
\implies \begin{cases} 
1. & B(x, r) = x + B(0, r) \quad \text{and} \\
2. & B(0, r) = r B(0, 1) \quad \text{and} \\
3. & B(x, r) \text{ is convex} \quad \text{and} \\
4. & x \in B(0, r) \iff -x \in B(0, r) \quad \text{(symmetric)} 
\end{cases}
\]

PROOF:

1. Proof that \(d(x + z, y + vz) = d(x, y)\) (invariant):

   \[
   \begin{align*}
   d(x + z, y + vz) &= \|(y + vz) - (x + z)\| \\
   &= \|y - x\| \quad \text{by left hypothesis} \\
   &= d(x, y) \quad \text{by left hypothesis}
   \end{align*}
   \]

2. Proof that \(B(x, r) = x + B(0, r)\):

   \[
   B(x, r) = \{ y \in X | d(x, y) < r \} \quad \text{by definition of open ball } B
   \]

   \[
   = \{ y \in X | d(y - x, y - x) < r \} \quad \text{by right result 1.}
   \]

   \[
   = \{ y \in X | d(0, y - x) < r \}
   \]

   \[
   = \{ u + x \in X | d(0, u) < r \} \quad \text{let } u \triangleq y - x
   \]

   \[
   = x + \{ u \in X | d(0, u) < r \}
   \]

   \[
   = x + B(0, r) \quad \text{by definition of open ball } B
   \]

3. Proof that \(B(0, r) = r B(0, 1)\):

   \[
   B(0, r) = \{ y \in X | d(0, y) < r \} \quad \text{by definition of open ball } B
   \]

   \[
   = \left\{ y \in X \left| \frac{1}{r} d(0, y) < 1 \right\} \right.
   \]

   \[
   = \left\{ y \in X \left| \frac{1}{r} \| y - 0 \| < 1 \right\} \right. \quad \text{by left hypothesis}
   \]

   \[
   = \left\{ y \in X \left| \frac{1}{r} y - \frac{1}{r} 0 \right\| < 1 \right\} \quad \text{by homogeneous property of } \|\cdot\| \text{ page 149}
   \]

   \[
   = \left\{ y \in X \left| d\left(\frac{1}{r} 0, \frac{1}{r} y\right) < 1 \right\} \right. \quad \text{by left hypothesis}
   \]

   \[
   = \left\{ ru \in X | d(0, u) < 1 \right\} \quad \text{let } u \triangleq \frac{1}{r} y
   \]

   \[
   = r \left\{ u \in X | d(0, u) < 1 \right\}
   \]

   \[
   = r B(0, 1) \quad \text{by definition of open ball } B
   \]

4. Proof that \(B(p, r)\) is convex:

   We must prove that for any pair of points \(x\) and \(y\) in the open ball \(B(p, r)\), any point \(\lambda x + (1 - \lambda)y\) is also in the open ball. That is, the distance from any point \(\lambda x + (1 - \lambda)y\) to the ball’s center \(p\) must be less

¹⁴\text{Giles (2000) page 2 (1.2 Remarks), }\text{Giles (1987) pages 22–26 (2.4 Theorem, 2.11 Theorem)}
than \( r \).

\[
d(p, \lambda x + (1 - \lambda)y) = \|p - \lambda x - (1 - \lambda)y\| \]
by left hypothesis

\[
= \left\| \frac{\lambda p + (1 - \lambda)p - \lambda x - (1 - \lambda)y}{p} \right\|
= \|\lambda p - \lambda x + (1 - \lambda)p - (1 - \lambda)y\|
\leq \|\lambda p - \lambda x\| + \|(1 - \lambda)p - (1 - \lambda)y\|
\]
by subadditivity property of \( \|\cdot\| \) page 149

\[
= |\lambda| \|p - x\| + |1 - \lambda| \|p - y\|
\]
by homogeneous property of \( \|\cdot\| \) page 149

\[
= \lambda \|p - x\| + (1 - \lambda) \|p - y\|
\]
because \( 0 \leq \lambda \leq 1 \)

\[
\leq \lambda r + (1 - \lambda)r
\]
because \( x, y \) are in the ball \( B(p, r) \)

\[
r
\]

5. Proof that \( x \in B(0, r) \iff -x \in B(0, r) \) (symmetric):

\[
x \in B(0, r) \iff x \in \{ y \in X | d(0, y) < r \}
\]
by definition of open ball \( B \)

\[
\iff x \in \{ y \in X | \|y - 0\| < r \}
\]
by left hypothesis

\[
\iff x \in \{ y \in X | \|y\| < r \}
\]
by definition of \( \|\cdot\| \)

\[
\iff x \in \{ y \in X | \|(-1)(-y)\| < r \}
\]
by homogeneous property of \( \|\cdot\| \) page 149

\[
\iff x \in \{ y \in X | \|-y\| < r \}
\]
by definition of \( \|\cdot\| \)

\[
\iff x \in \{ y \in X | \|-y - 0\| < r \}
\]
by left hypothesis

\[
\iff x \in \{ -u \in X | d(0, u) < r \}
\]
le t \( u \triangleq -y \)

\[
\iff x \in \{ u \in X | d(0, u) < r \}
\]
le t \( u \triangleq -y \)

\[
\iff -x \in B(0, r)
\]

\[
\iff -x \in B(0, r)
\]

\[

\Rightarrow
\]

Theorem D.4 (page 150) demonstrates that if a metric \( d \) in a metric space \( (X, +, \cdot, (F, +, \times), d) \) is generated by a norm, then the ball \( B(x, r) \) in that metric linear space is **convex**. However, the converse is not true. That is, it is possible for the balls in a metric space \( (Y, p) \) to be convex, but yet the metric \( p \) not be generated by a norm.

**Norms generated by metrics**

Every normed linear space is also a metric linear space (Theorem D.3 page 149). However, the converse is not true—not every metric linear space is a **normed linear space**. A characterization of metric linear spaces that are normed linear spaces is provided by Theorem D.5 (page 152).

**Lemma D.1.** 15 \( \{ X, +, \cdot, (F, +, \times), d \} \) be a metric linear space. Let \( \|x\| \triangleq d(x, 0) \) \( \forall x \in X \).

\[
d(x + z, y + z) = d(x, y) \quad \forall x, y, z \in X
\]

**PROOF:**

---

15 Oikhberg and Rosenthal (2007) page 599
1. Proof that \( \|x\| = \|-x\| \):
\[
\|x\| = d(x, 0) \\
d = d(x - x, 0 - x) \\
= d(0, -x) \\
\|\|-x\| \\
by definition of \( \|\cdot\| \) \]
by translation invariance hypothesis
by definition of \( \|\cdot\| \)
by property of metrics

2a. Proof that \( \|x\| = 0 \implies x = 0 \):
\[
0 = \|x\| \\
d(x, 0) \\
= d(0, 0) \\
\|x\| = 0 \\
by left hypothesis \]
by definition of \( \|\cdot\| \) \]
by right hypothesis \]
by property of metrics

2b. Proof that \( \|x\| = 0 \iff x = 0 \):
\[
\|x\| = d(x, 0) \\
= d(0, 0) \\
= 0 \\
\|x\| = 0 \\
by definition of \( \|\cdot\| \) \]
by left hypothesis \]
by property of metrics

3. Proof that \( \|x + y\| \leq \|x\| + \|y\| \):
\[
\|x + y\| = d(x + y, 0) \\
= d(x + y - y, 0 - y) \\
\leq d(x, 0) + d(0, y) \\
\|x\| + \|y\| \\
by definition of \( \|\cdot\| \) \]
by translation invariance hypothesis \]
by property of metrics \]
by property of metrics \]
by definition of \( \|\cdot\| \) \]

**Theorem D.5.** \(^{16}\) Let \((X, +, \cdot, (F, \oplus, \odot))\) be a LINEAR SPACE. Let \(d(x, y) \triangleq \|x - y\| \forall x, y \in X\).

\[1. \quad d(x + z, y + z) = d(x, y) \quad \forall x, y, z \in X \text{ (translation invariant)} \quad \text{and} \]
\[2. \quad d(\alpha x, \alpha y) = |\alpha| d(x, y) \quad \forall x, y \in X, \alpha \in F \text{ (homogeneous)} \]
\[\iff \|\cdot\| \text{ is a NORM} \]

**Proof:**

1. Proof of \( \iff \) assertion:

   (a) Proof that \( \|\cdot\| \) is strictly positive: This follows directly from the definition of \( d \).

   (b) Proof that \( \|\cdot\| \) is nondegenerate: This follows directly from Lemma D.1 (page 151).

   (c) Proof that \( \|\cdot\| \) is homogeneous: This follows from the second left hypothesis.

   (d) Proof that \( \|\cdot\| \) satisfies the triangle-inequality: This follows directly from Lemma D.1 (page 151).

2. Proof of \( \iff \) assertion:
\[
d(x + z, y + z) = \|(x + z) - (y + z)\| \\
= \|x - y\| \\
= d(x, y) \\
by definition of \( d \) \]
by definition of \( d \)
\[
d(\alpha x, \alpha y) = \|\alpha x - \alpha y\| \\
= \|\alpha(x - y)\| \\
= |\alpha| \|x - y\| \\
by definition of \( \|\cdot\| \) page 149 \]
by definition of \( d \)

\(^{16}\) Bollobás (1999), page 21
D.2 Metric spaces

"The Epicureans are wont to ridicule this theorem, saying it is evident even to an ass and needs no proof; it is as much the mark of an ignorant man, they say, to require persuasion of evident truths as to believe what is obscure without question... That the present theorem is known to an ass they make out from the observation that, if straw is placed at one extremity of the sides, an ass in quest of provender will make his way along the one side and not by way of the two others."

Proclus Lycaeus (412 – 485 AD), Greek philosopher, commenting on the Epicureans view of the triangle inequality property.

D.2.1 Algebraic structure

Definition D.6.  A function \( d \in \mathbb{R}^{+} \times \mathbb{R}^{+} \) (Definition 1.6 page 6) is a quasi-metric on a set \( X \) if

1. \( d(x, y) \geq 0 \) \( \forall x, y \in X \) (non-negative) and
2. \( d(x, y) = 0 \iff x = y \) \( \forall x, y \in X \) (non-degenerate) and
3. \( d(x, y) \leq d(x, z) + d(z, y) \) \( \forall x, y, z \in X \) (subadditive/oriented triangle inequality).

The pair \((X, d)\) is a quasi-metric space if \( d \) is a quasi-metric on \( X \). A quasi-metric is also called an asymmetric metric and a directed metric.

Definition D.7.  Let \( X \) be a set and \( \mathbb{R}^{+} \) the set of non-negative real numbers.

A function \( d \in \mathbb{R}^{+} \times \mathbb{R}^{+} \) is a metric on \( X \) if

1. \( d(x, y) \geq 0 \) \( \forall x, y \in X \) (non-negative) and
2. \( d(x, y) = d(y, x) \) \( \forall x, y \in X \) (symmetric) and
3. \( d(x, y) \leq d(x, z) + d(z, y) \) \( \forall x, y, z \in X \) (subadditive/triangle inequality).

A metric space is the pair \((X, d)\). A metric is also called a distance function.

Actually, it is possible to significantly simplify the definition of a metric to an equivalent statement requiring only half as many conditions. These equivalent conditions (a “characterization”) are stated in Theorem D.6 (next).

Theorem D.6 (metric characterization).  Let \( d \) be a function in \((\mathbb{R}^{+})^{X \times X}\).

\( d(x, y) \) is a metric \iff \begin{align*} 1. \quad & d(x, y) = 0 \iff x = y \quad \forall x, y \in X \quad \text{and} \quad \text{and} \\ 2. \quad & d(x, y) \leq d(z, x) + d(z, y) \quad \forall x, y, z \in X \end{align*}

\( \Box \) Proof:

1. Proof that \([d(x, y) \text{ is a metric}] \implies [(1) \text{ and } (2)]:\)

\( ^{17} \) Lycaeus (circa 450), page 251


\( ^{19} \) Dieudonné (1969), page 28, Copson (1968), page 21, Hausdorff (1937) page 109, Fréchet (1928), Fréchet (1906) page 30

\( ^{20} \) Euclid (circa 300BC) (Book I Proposition 20)

\( ^{21} \) Busemann (1955) page 3, Michel and Herget (1993), page 264, Giles (1987), page 18
1a. Proof that \( d(x, y) = 0 \iff x = y \): by left hypothesis 2 (\( d(x, y) \) is nondegenerate)

1b. Proof that \( d(x, y) \leq d(z, x) + d(z, y) \):

\[
d(x, y) \leq d(z, x) + d(z, y)
\]

by right hypothesis 4 (triangle inequality)

\[
d(z, x) + d(z, y)
\]

by right hypothesis 3 (commutative)

2. Proof that \( \{d(x, y) \) is a metric \( \iff \) (1) and (2)]:

2a. Proof that \( d(x, y) \geq 0 \):

\[
0 = \frac{1}{2} \cdot 0
\]

\[
= \frac{1}{2} d(y, y)
\]

by right hypothesis 1

\[
= \frac{1}{2} d(y, z)_{z=y}
\]

\[
\leq \frac{1}{2} [d(x, y) + d(x, z)]_{z=y}
\]

by right hypothesis 2

\[
= \frac{1}{2} [d(x, y) + d(x, y)]
\]

\[
= d(x, y)
\]

2b. Proof that \( d(x, y) = 0 \iff x = y \): by right hypothesis 1

2c. Proof that \( d(x, y) = d(y, x) \):

\[
d(x, y)_{z=x} \leq [d(z, x) + d(z, y)]_{z=x}
\]

by right hypothesis 2

\[
d(y, x)_{z=x}
\]

by right hypothesis 1

2d. Proof that \( d(x, y) \leq d(x, z) + d(z, y) \):

\[
d(x, y) \leq d(z, x) + d(z, y)
\]

by right hypothesis 2

\[
= d(x, z) + d(z, y)
\]

by result 2c

The triangle inequality property stated in the definition of metrics (Definition D.7 page 153) axiomatically endows a metric with an upper bound. Lemma D.2 (next) demonstrates that there is a complementary lower bound similar in form to the triangle-inequality upper bound.

**Lemma D.2.** Let \( (X, d) \) be a metric space (Definition D.7 page 153). Let \(|\cdot|\) be the absolute value function (Definition 1.28 page 11).

1. \(|d(x, p) - d(p, y)| \leq d(x, y) \quad \forall x, y, p \in X\)

2. \(d(x, p) - d(p, y) \leq d(x, y) \quad \forall x, y, p \in X\)

**Proof:**

\(^{22}\) Dieudonné (1969), page 28, Michel and Herget (1993), page 266
1. Proof that \( |d(x, p) - d(p, y)| \leq d(x, y) \):

\[
|d(x, p) - d(p, y)| \leq |d(x, y) + d(y, p) - d(p, y)| \quad \text{by subadditive property (Definition D.7 page 153)}
\]

\[
= |d(x, y) + d(p, y) - d(p, y)|
\]

\[
= |d(x, y) + 0|
\]

\[
= d(x, y)
\]

by non-negative property of metrics (Definition D.7 page 153)

2. Proof that \( d(x, p) \geq d(p, y) \implies d(x, p) - d(p, y) \leq d(x, y) \):

\[
d(x, p) - d(p, y) = |d(x, p) - d(p, y)| \quad \text{by left hypothesis and definition of } |\cdot|
\]

\[
\leq d(x, y) \quad \text{by item (1)}
\]

3. Proof that \( d(x, p) \leq d(p, y) \implies d(x, p) - d(p, y) \leq d(x, y) \):

\[
|d(x, p) - d(p, y)| \leq 0 \quad \text{by left hypothesis}
\]

\[
\leq d(x, y) \quad \text{by non-negative property of metrics (Definition D.7 page 153)}
\]

The triangle inequality property stated in the definition of metrics (Definition D.7 page 153) can be extended from two to any finite number of metrics (next).

**Proposition D.1.**\(^{23}\) Let \((X, d)\) be a metric space (Definition D.7 page 153) and \(\{x_n \in X\}_{i=1}^N\) an \(N\)-tuple (Definition 1.11 page 6) on \(X\).

\[
d(x_1, x_N) \leq \sum_{n=1}^{N-1} d(x_n, x_{n+1}) \quad \forall N \in \mathbb{N} \setminus \{1\}
\]

**Proof:** Proof by induction:

Proof that the \(\{N = 2\}\) case is true:

\[
d(x_1, x_2) \leq \sum_{n=1}^{2-1} d(x_n, x_{n+1})
\]

Proof for that the \(\{N\} \implies \{N + 1\}\) case:

\[
d(x_1, x_{N+1}) \leq d(x_1, x_N) + d(x_N, x_{N+1}) \quad \text{by subadditive property (Definition D.7 page 153)}
\]

\[
\leq \left( \sum_{n=1}^{N-1} d(x_n, x_{n+1}) \right) + d(x_N, x_{N+1}) \quad \text{by } \{N \text{ case} \} \text{ hypothesis}
\]

\[
= \sum_{n=1}^{N} d(x_n, x_{n+1})
\]

**Definition D.8.**\(^{24}\) Let \(X\) be a set and \(d \in \mathbb{R}^{X \times X}\). The function \(d\) is the discrete metric on \(\mathbb{R}^{X \times X}\) if

\[
d(x, y) \triangleq \begin{cases} 0 & \text{if } x = y \\ 1 & \text{otherwise} \end{cases} \quad \forall x, y \in X
\]

\(\text{\^{23}}\) Dieudonné (1969), page 28

\(\text{\^{24}}\) Rosenlicht (1968) page 37

\(\text{\^{24}}\) Busemann (1955) page 4 (Comments on the axioms), \(\text{\^{24}}\) Giles (1987), page 13, \(\text{\^{24}}\) Copson (1968), page 24, \(\text{\^{24}}\) Khamsi and Kirk (2001) page 19 (Example 2.1)
Definition D.9 (usual metric). Let $|\cdot| \in \mathbb{R}^{+R}$ be an absolute value function on a ring $R$. The function $d(x, y) \triangleq |x - y|$ is a metric on $\mathbb{R}$, called the usual metric.

Definition D.10. Let $X$ be a set and $d \in \mathbb{R}^{N\times R}$. The function $d$ is the Euclidean metric on $\mathbb{R}^N$ if

$$d((x_1, x_2, \ldots, x_N), (y_1, y_2, \ldots, y_N)) \triangleq \sqrt{\sum_{n=1}^{N}(x_n - y_n)^2} \quad \forall (x_1, x_2, \ldots, x_N), (y_1, y_2, \ldots, y_N) \in \mathbb{R}^N$$

D.2.2 Metric preserving functions

Definition D.11. Let $\mathcal{M}$ be the set of all METRIC SPACES (Definition D.7 page 153) on a set $X$. $\phi \in \mathbb{R}^{+R\times R}$ is a metric preserving function if $d(x, y) \triangleq \phi \star p(x, y)$ is a metric on $X$ for all $(X, p) \in \mathcal{M}$

Theorem D.7 (necessary conditions). Let $\mathcal{R}\phi$ be the range of a function $\phi$.

$$\begin{align*}
\{ \phi & \text{ is a } \text{METRIC PRESERVING FUNCTION} \} \\
\quad \{ \text{Definition D.11 page 156} \} & \implies \\
& \{ 1. \phi^{-1}(0) = \{0\} \quad \text{and} \\
& 2. \mathcal{R}\phi \subseteq \mathbb{R}^+ \quad \text{and} \\
& 3. \phi(x + y) \leq \phi(x) + \phi(y) \quad (\phi \text{ is SUBADDITIONAL}) \}
\end{align*}$$

Proof:

1. Proof that $\phi$ is a metric preserving function $\implies \phi^{-1}(0) = \{0\}$:

   (a) Suppose that the statement is not true and $\phi^{-1}(0) = \{0, a\}$.
   (b) Then $\phi(a) = 0$ and for some $x, y$ such that $x \neq y$ and $d(x, y) = a$ we have

   $$\phi \star d(x, y) = \phi(a)$$

   $$= 0$$

   $$\implies \phi \star d \text{ is not a metric}$$

   $$\implies \phi \text{ is not a metric preserving function}$$

   (c) But this contradicts the original hypothesis, and so it must be that $\phi^{-1}(0) = \{0\}$.

2. Proof that $\mathcal{R}\phi \subseteq \mathbb{R}^+$:

   $$\mathcal{R}\phi \star d \subseteq \mathcal{R}d$$

   $$\subseteq \mathbb{R}^+$$

3. Proof that $\phi$ is a metric preserving function $\implies \phi$ is subadditive:

   (a) For $\phi$ to be a metric preserving function, by definition it must work with all metric spaces.
   (b) So to develop necessary conditions, we can pick any metric space we want (because it is necessary that $\phi$ preserves it as a metric space).
   (c) For this proof we choose the metric space $(\mathbb{R}, d)$ where $d(x, y) \triangleq |x - y|$ for all $x, y \in \mathbb{R}^+$:

   $$\phi(x) + \phi(y) = \phi(|(x + y) - x|) + \phi(|x - 0|)$$

   by definition of $|\cdot|$ = $(\phi \star d)(x + y, x) + (\phi \star d)(x, 0)$

   by definition of $d$ \geq (\phi \star d)(x + y, 0, 0)

   by left hypothesis and Definition D.7 page 153

   = $\phi(|(x + y) - 0|)$

   by definition of $d$

   = $\phi(x + y)$

   because $x, y \in \mathbb{R}^+$

---

25 Davis (2005) page 16
26 Vallin (1999), page 849 (Definition 1.1), Corazza (1999), page 309, Deza and Deza (2009) page 80
27 Corazza (1999), page 310 (Proposition 2.1), Deza and Deza (2009) page 80
Theorem D.8 (next theorem) presents some sufficient conditions for a function to be metric-preserving.

**Theorem D.8 (sufficient conditions).** Let \( \phi \) be a function in \( \mathbb{R}^R \).

\[
\begin{align*}
1. & \quad x \geq y \implies \phi(x) \geq \phi(y) \quad \forall x, y \in \mathbb{R}^R \quad \text{(isotone)} \\
2. & \quad \phi(0) = 0 \\
3. & \quad \phi(x + y) \leq \phi(x) + \phi(y) \quad \forall x, y \in \mathbb{R}^R \quad \text{(subadditive)}
\end{align*}
\]

\[
\implies \begin{cases} 
\phi \text{ is a metric preserving function} \\
\text{(Definition D.11 page 156).}
\end{cases}
\]

**Proof:**

1. Proof that \( \phi \circ d(x, y) = 0 \implies x = y: \)

\[
\phi \circ d(x, y) = 0 \implies d(x, y) = 0 \\
\implies x = y \quad \text{by nondegenerate property page 153}
\]

2. Proof that \( \phi \circ d(x, y) = 0 \iff x = y: \)

\[
\phi \circ d(x, y) = \phi \circ d(x, x) \\
= \phi(0) \\
= 0 \quad \text{by nondegenerate property page 153}
\]

3. Proof that \( \phi \circ d(x, y) \leq \phi \circ d(z, x) + \phi \circ d(z, y): \)

\[
\phi \circ d(x, y) \leq \phi \circ d(z, z) + \phi \circ d(z, y) \quad \text{by symmetric property of} \ d \text{ page 153}
\]

\[
\leq \phi \circ d(z, x) + \phi \circ d(z, y) \quad \text{by} \ \phi \ \text{hypothesis 3}
\]

\[
\implies \begin{cases} 
\phi \text{ is a metric preserving function} \\
\text{(Definition D.11 page 156).}
\end{cases}
\]

\[
\begin{array}{c}
\text{Figure D.1: metric preserving functions}
\end{array}
\]

**Example D.3 (\( \alpha \)-scaled metric/dilated metric).** Let \((X, d)\) be a metric space (Definition D.7 page 153). \( \phi(x) \triangleq \alpha x, \ \alpha \in \mathbb{R}^+ \) is a metric preserving function (Figure D.1 page 157 (A)).

**Proof:** The proofs for Example D.3–Example D.8 (page 158) follow from Theorem D.8 (page 157).
Example D.4 (power transform metric/snowflake transform metric). Let \((X, d)\) be a metric space (Definition D.7 page 153). \(\phi(x) \triangleq x^\alpha, \alpha \in (0 : 1)\), is a metric preserving function (see Figure D.1 page 157 (B)).

Example D.5 (\(\alpha\)-truncated metric/radar screen metric). Let \((X, d)\) be a metric space (Definition D.7 page 153). \(\phi(x) \triangleq \min \{\alpha, x\}, \alpha \in \mathbb{R}^+\) is a metric preserving function (see Figure D.1 page 157 (C)).

\[\phi(x) = \begin{cases} 0 & \text{for } x \leq 0, \\ 1 & \text{otherwise} \end{cases}\]

Example D.6 (bounded metric). Let \((X, d)\) be a metric space (Definition D.7 page 153). \(\phi(x) \triangleq \frac{x}{1 + x}\) is a metric preserving function (see Figure D.1 page 157 (D)).

\[\phi(x) = \begin{cases} \frac{x}{1 + x} & \text{is a metric preserving function (see Figure D.1 page 157 (D)).} \\ \end{cases}\]

Example D.7 (discrete metric preserving function). Let \(\phi\) be a function in \(\mathbb{R}^R\).

\[\phi(x) = \begin{cases} 0 & \text{for } x \leq 0, \\ 1 & \text{otherwise} \end{cases}\]

Example D.8. Let \(\phi\) be a function in \(\mathbb{R}^R\).

\[\phi(x) = \begin{cases} x & \text{for } 0 \leq x < 1, \\ 1 & \text{for } 1 \leq x \leq 2, \\ x - 1 & \text{for } 2 < x < 3, \\ 2 & \text{for } x \geq 3 \end{cases}\]

Example D.9. Let \(\phi\) be a function in \(\mathbb{R}^R\).

\[\phi(x) = \begin{cases} 0 & \text{for } x = 0, \\ \frac{1}{1 + \frac{x}{x+1}} & \text{for } x > 0 \end{cases}\]

Example D.10. Let \(\phi\) be a function in \(\mathbb{R}^R\).

\[\phi(x) = \begin{cases} x & \text{for } x \leq 2, \\ 1 + \frac{1}{x-1} & \text{for } x > 2 \end{cases}\]

Example D.11. Let \(\phi\) be a function in \(\mathbb{R}^R\).

\[\phi(x) = \begin{cases} x & \text{for } 0 \leq x \leq 2, \\ x + 4 & \text{for } 2 < x < 3, \\ 1 & \text{for } x \geq 3 \end{cases}\]

\[\phi(x) = \begin{cases} 0 & \text{for } x = 0, \\ \frac{1}{1 + \frac{x}{x+1}} & \text{for } x > 0 \end{cases}\]

\[\phi(x) = \begin{cases} x & \text{for } x \leq 2, \\ 1 + \frac{1}{x-1} & \text{for } x > 2 \end{cases}\]

\[\phi(x) = \begin{cases} x & \text{for } 0 \leq x \leq 2, \\ x + 4 & \text{for } 2 < x < 3, \\ 1 & \text{for } x \geq 3 \end{cases}\]

---

33 Corazza (1999), page 311
34 Greenhoe (2015), pages 10–11 (Theorem 4.16)
D.2.3 Product metrics

Theorem D.9 (Fréchet product metric). \(^{36}\) Let \(X\) be a set.

\[
\begin{align*}
1. \{p_n\} \text{ are metrics on } X \quad \text{and} \\
2. \alpha_n \geq 0 \quad \forall n = 1, 2, \ldots, N \quad \text{and} \\
3. \max \left\{ \alpha_n \left| n = 1, 2, \ldots, N \right. \right\} > 0 \\
\end{align*}
\]
\[
\implies \quad \begin{cases}
\text{d}(x, y) = \sum_{n=1}^{N} \alpha_n p_n(x, y) \\
\text{is a metric on } X
\end{cases}
\]

\(\Box\) Proof:

1. Proof that \(x = y \implies d(x, y) = 0:\)

\[
d(x, y) = \sum_{n=1}^{N} \alpha_n p_n(x, y) \quad \text{by definition of } d
\]
\[
= \sum_{n=1}^{N} \alpha_n p_n(x, x) \quad \text{by left hypothesis}
\]
\[
= \sum_{n=1}^{N} 0 \quad \text{by non-degenerate property of metrics (Definition D.7 page 153)}
\]
\[
= 0
\]

2. Proof that \(x = y \iff d(x, y) = 0:\)

\[
0 = d(x, y) \quad \text{by right hypothesis}
\]
\[
= \sum_{n=1}^{N} \alpha_n p_n(x, y) \quad \text{by definition of } d
\]
\[
\implies p_n(x, y) = 0 \quad \forall x, y \in X \quad \text{by metric properties page 153}
\]
\[
\implies x = y \quad \forall x, y \in X \quad \text{by non-degenerate property of metrics page 153}
\]

3. Proof that \(d(x, y) \leq d(x, z) + d(z, y):\)

\[
d(x, y) = \sum_{n=1}^{N} \alpha_n p_n(x, y) \quad \text{by definition of } d
\]
\[
\leq \sum_{n=1}^{N} \alpha_n \left[ p_n(x, z) + p_n(z, y) \right] \quad \text{by subadditive property (Definition D.7 page 153)}
\]
\[
= \sum_{n=1}^{N} \alpha_n \left[ p_n(x, z) + p_n(z, y) \right] \quad \text{by symmetry property (Definition D.7 page 153)}
\]
\[
= \sum_{n=1}^{N} \alpha_n p_n(z, x) + \sum_{n=1}^{N} \alpha_n p_n(z, y) \quad \text{by subadditive property (Definition D.7 page 153)}
\]
\[
= \text{d}(z, x) + \text{d}(z, y) \quad \text{by definition of } d
\]

\(\Box\)

Theorem D.10 (Power mean metrics). Let \(X\) be a set. Let \(\langle x_n \in X \rangle_1^N\) and \(\langle y_n \in X \rangle_1^N\) be \(N\)-tuples on \(X\).

\[
\begin{align*}
1. \quad p \text{ is a metric on } X \quad \text{and} \\
2. \quad \sum_{n=1}^{N} \lambda_n = 1
\end{align*}
\]
\[
\implies \quad \left\{ \begin{array}{c}
\text{d}(\langle x_n \rangle_1^N, \langle y_n \rangle_1^N) \triangleq \left( \sum_{n=1}^{N} \lambda_n p(x_n, y_n) \right)^{\frac{1}{r}} \quad \forall r \in [1: \infty]
\end{array} \right\}
\]

Moreover, if \(r = \infty\), then \(d(\langle x_n \rangle_1^N, \langle y_n \rangle_1^N) = \max_{n=1, \ldots, N} p(x_n, y_n).\)

Proof:

1. Proof that $\|x_n\| = \|y_n\| \implies d(\|x_n\|, \|y_n\|) = 0$ for $r \in [1, \infty)$:

$$d(\|x_n\|, \|y_n\|) \triangleq \left( \sum_{n=1}^{N} \lambda_n p(r(x_n, y_n)) \right)^{\frac{1}{r}}$$

by definition of $d$

$$= \left( \sum_{n=1}^{N} \lambda_n p(r(x_n, y_n)) \right)^{\frac{1}{r}}$$

by $\|x_n\| = \|y_n\|$ hypothesis

$$= \left( \sum_{n=1}^{N} 0 \right)^{\frac{1}{r}}$$

because $p$ is non-degenerate

$$= 0$$

2. Proof that $\|x_n\| = \|y_n\| \iff d(\|x_n\|, \|y_n\|) = 0$ for $r \in [1, \infty)$:

$$0 = d(\|x_n\|, \|y_n\|)$$

by $d(\|x_n\|, \|y_n\|) = 0$ hypothesis

$$\triangleq \left( \sum_{n=1}^{N} \lambda_n p(r(x_n, y_n)) \right)^{\frac{1}{r}}$$

by definition of $d$

$$\implies (p(z_n, y_n))^{\frac{1}{r}} = 0 \text{ for } n = 1, 2, \ldots, N$$

because $p$ is non-negative

$$\implies \|x_n\| = \|y_n\|$$

because $p$ is non-degenerate

3. Proof that $d$ satisfies the triangle inequality property for $r = 1$:

$$d(\|x_n\|, \|y_n\|) \triangleq \left( \sum_{n=1}^{N} \lambda_n p(x_n, y_n) \right)$$

by definition of $d$

$$= \sum_{n=1}^{N} \lambda_n p(x_n, y_n)$$

by $r = 1$ hypothesis

$$\leq \sum_{n=1}^{N} \lambda_n [p(z_n, x_n) + p(z_n, y_n)]$$

by triangle inequality

$$= \sum_{n=1}^{N} \lambda_n p(z_n, x_n) + \sum_{n=1}^{N} \lambda_n p(z_n, y_n)$$

$$= \left( \sum_{n=1}^{N} \lambda_n p'(z_n, x_n) \right)^{\frac{1}{r}} + \left( \sum_{n=1}^{N} \lambda_n p'(z_n, y_n) \right)^{\frac{1}{r}}$$

by $r = 1$ hypothesis

$$\triangleq d(\|z_n\|, \|x_n\|) + d(\|z_n\|, \|y_n\|)$$

by definition of $d$

4. Proof that $d$ satisfies the triangle inequality property for $r \in (1, \infty)$:

$$d(\|x_n\|, \|y_n\|)$$

$$\triangleq \left( \sum_{n=1}^{N} \lambda_n p'(x_n, y_n) \right)^{\frac{1}{r}}$$

by definition of $d$

$$\leq \left( \sum_{n=1}^{N} \lambda_n \left[ p(z_n, x_n) + p(z_n, y_n) \right]^{r} \right)^{\frac{1}{r}}$$

by subadditive property (Definition D.7 page 153)
\[
\begin{align*}
&= \left( \sum_{n=1}^{N} \left[ \lambda_n^\frac{1}{p}(z_n, x_n) + \lambda_n^\frac{1}{q}(z_n, y_n) \right]^{\frac{1}{p}} \right)^{\frac{1}{\frac{1}{p}}} \\
&\leq \left( \sum_{n=1}^{N} \left[ \lambda_n^\frac{1}{p}(z_n, x_n) \right]^{\frac{1}{p}} \right) + \left( \sum_{n=1}^{N} \left[ \lambda_n^\frac{1}{q}(z_n, y_n) \right]^{\frac{1}{q}} \right)
\end{align*}
\]

by subadditive property (Definition D.7 page 153)

\[
\leq \left( \sum_{n=1}^{N} \lambda_n^\frac{1}{p}(z_n, x_n) \right)^{\frac{1}{p}} + \left( \sum_{n=1}^{N} \lambda_n^\frac{1}{q}(z_n, y_n) \right)^{\frac{1}{q}}
\]

by Minkowski's inequality

\[
\leq d(\langle z_n \rangle, \langle x_n \rangle) + d(\langle z_n \rangle, \langle y_n \rangle)
\]

by definition of \(d\)

5. Proof for the \(r = \infty\) case:

(a) Proof that \(d(\langle x_n \rangle, \langle y_n \rangle) = \max \{x_n\}\) by Theorem D.14 page 164

(b) Proof that \(\langle x_n \rangle = \langle y_n \rangle \implies d(\langle x_n \rangle, \langle y_n \rangle) = 0:\)

\[
d(\langle x_n \rangle, \langle y_n \rangle) \triangleq \max \{x_n, y_n\}
\]

by definition of \(d\)

\[
= \max \{p(x_n, y_n) | n = 1, 2, \ldots, N\}
\]

by \(\langle x_n \rangle = \langle y_n \rangle\) hypothesis

\[
= 0
\]

because \(p\) is nondegenerate

(c) Proof that \(\langle x_n \rangle = \langle y_n \rangle \iff d(\langle x_n \rangle, \langle y_n \rangle) = 0:\)

\[
0 = d(\langle x_n \rangle, \langle y_n \rangle)
\]

by \(d(\langle x_n \rangle, \langle y_n \rangle) = 0\) hypothesis

\[
\triangleq \max \{p(x_n, y_n) | n = 1, 2, \ldots, N\}
\]

by definition of \(d\)

\[
\implies p(x_n, y_n) = 0 \text{ for } n = 1, 2, \ldots, N
\]

because \(p\) is nondegenerate

(d) Proof that \(d\) satisfies the triangle inequality property:

\[
d(\langle x_n \rangle, \langle y_n \rangle)
\]

\[
\triangleq \max \{p(x_n, y_n) | n = 1, 2, \ldots, N\}
\]

by definition of \(d\)

\[
\leq \max \{p(x_n, z_n) + p(z_n, y_n) | n = 1, 2, \ldots, N\}
\]

by subadditive property

\[
\leq \max \{p(x_n, z_n) | n = 1, 2, \ldots, N\} + \max \{p(z_n, y_n) | n = 1, 2, \ldots, N\}
\]

by non-negative property

\[
= \max \{p(x_n, z_n) | n = 1, 2, \ldots, N\} + \max \{p(z_n, y_n) | n = 1, 2, \ldots, N\}
\]

by symmetry property

\[
\triangleq d(\langle z_n \rangle, \langle x_n \rangle) + d(\langle z_n \rangle, \langle y_n \rangle)
\]

by definition of \(d\)

6. And so by Theorem D.6 (page 153), \(d\) is a metric for \(r \in [1 : \infty]\).

\[\square\]

D.3 Sums

"I think that it was Harald Bohr who remarked to me that "all analysts spend half their time hunting through the literature for inequalities which they want to use and cannot prove.""

G.H. Hardy (1877–1947) in his “Presidential Address” to the London Mathematical Society on November 8, 1928, about a remark that he thought to be from Harald Bohr (1887–1951), Danish mathematician and pictured to the left. \(^{37}\)

\(^{37}\) quote: \(\equiv\) Hardy (1929), page 64
image: [http://www-history.mcs.st-andrews.ac.uk/PictDisplay/Bohr_Harald.html](http://www-history.mcs.st-andrews.ac.uk/PictDisplay/Bohr_Harald.html)
D.3.1 Summation

Definition D.12.\(^{38}\) Let \(+\) be an addition operator on a tuple \(\langle x_n \rangle^N_m\).

The summation of \(\langle x_n \rangle\) from index \(m\) to index \(N\) with respect to \(+\) is

\[
\sum_{n=m}^{N} x_n \triangleq \begin{cases} 
0 & \text{for } N < m \\
\left( \sum_{n=m}^{N-1} x_n \right) + x_N & \text{for } N \geq m
\end{cases}
\]

Theorem D.11 (Generalized associative property).\(^{39}\) Let \(+\) be an addition operator on a tuple \(\langle x_n \rangle^N_m\).

\[\sum_{n=m}^{N} x_n + \left( \sum_{n=m}^{L} x_n + \sum_{n=L+1}^{N} x_n \right) = \left( \sum_{n=m}^{L} x_n + \sum_{n=L+1}^{M} x_n \right) + \sum_{n=M+1}^{N} x_n \quad \text{for } m < L < M \leq N\]

\[\sum_{n=m}^{N} \text{ is ASSOCIATIVE}\]

Proof:

1. Proof for \(N < m\) case: \(\sum_{n=m}^{N} x_n = 0\).

2. Proof for \(N = m\) case: \(\sum_{n=m}^{m} x_n = \left( \sum_{n=m}^{m-1} x_n \right) + x_m = 0 + x_m = x_m\).

3. Proof for \(N = m + 1\) case: \(\sum_{n=m}^{m+1} x_n = \left( \sum_{n=m}^{m} x_n \right) + x_{m+1} = x_m + x_{m+1}\).

4. Proof for \(N = m + 2\) case:

\[
\sum_{n=m}^{m+2} x_n = \left( \sum_{n=m}^{m+1} x_n \right) + x_{m+2} \quad \text{by Definition D.12 page 162}
\]

\[
= (x_m + x_{m+1}) + x_{m+2} \quad \text{by item (3)}
\]

\[
= x_m + (x_{m+1} + x_{m+2}) \quad \text{by left hypothesis}
\]

5. Proof that \(N\) case \(\implies\) \(N + 1\) case:

\[
\sum_{n=m}^{N+1} x_n = \left( \sum_{n=m}^{N} x_n \right) + x_{N+1} \quad \text{by Definition D.12 page 162}
\]

\[
\text{associative}
\]

\[
= \left( \sum_{n=m}^{L} x_n + \left( \sum_{n=L+1}^{M} x_n + \sum_{n=M+1}^{N} x_n \right) \right) + x_{N+1}
\]

\[
= \left( \sum_{n=m}^{L} x_n + \sum_{n=L+1}^{M} x_n + \sum_{n=M+1}^{N} x_n \right) + x_{N+1}
\]

\(^{38}\) Berberian (1961) page 8 (Definition I.3.1),\(^{39}\) Fourier (1820) page 280 ("\(\sum\" notation)

\(^{39}\) Berberian (1961) pages 9–10 (Theorem I.3.1)
\[
\begin{aligned}
&= \left( \sum_{n=m}^{L} x_n + \sum_{n=L+1}^{M} x_n \right) + \left( \sum_{n=M+1}^{N} x_n + x_{N+1} \right) \\
&= \left( \sum_{n=m}^{L} x_n + \sum_{n=L+1}^{M} x_n \right) + \left( \sum_{n=M+1}^{N} x_n \right)
\end{aligned}
\]

\[\square\]

D.3.2 Convexity

**Definition D.13.** \(^{40}\) A function \( f \in \mathbb{R}^2 \) is convex if
\[
f(\lambda x + [1 - \lambda]y) \leq \lambda f(x) + (1 - \lambda) f(y) \quad \forall x, y \in \mathbb{R} \text{ and } \forall \lambda \in (0 : 1)
\]
A function \( g \in \mathbb{R}^2 \) is strictly convex if
\[
g(\lambda x + [1 - \lambda]y) < \lambda g(x) + (1 - \lambda) g(y) \quad \forall x, y \in D, x \neq y, \text{ and } \forall \lambda \in (0 : 1)
\]
A function \( f \in \mathbb{R}^2 \) is concave if \(-f\) is convex.
A function \( f \in \mathbb{R}^2 \) is affine if \( f \) is convex and concave.

**Theorem D.12** (Jensen’s Inequality). \(^{41}\) Let \( f \in \mathbb{R}^2 \) be a function.
\[
\left\{ \begin{array}{l}
\text{1. } f \text{ is convex and } \\
\text{2. } \sum_{n=1}^{N} \lambda_n = 1
\end{array} \right\} \implies \left\{ \begin{array}{l}
f \left( \sum_{n=1}^{N} \lambda_n x_n \right) \leq \sum_{n=1}^{N} \lambda_n f(x_n) \quad \forall x_n \in D, N \in \mathbb{N}
\end{array} \right\}
\]

D.3.3 Power means

**Definition D.14.** \(^{42}\) The \( \langle \lambda_n \rangle_1^N \) weighted \( \phi \)-mean of a tuple \( \langle x_n \rangle_1^N \) is defined as
\[
M_\phi(\langle x_n \rangle) \triangleq \phi^{-1} \left( \sum_{n=1}^{N} \lambda_n \phi(x_n) \right)
\]
where \( \phi \) is a continuous and strictly monotonically increasing function in \( \mathbb{R}^\mathbb{R} \)
and \( \langle \lambda_n \rangle_1^N \) is a sequence of weights for which \( \sum_{n=1}^{N} \lambda_n = 1 \).

**Lemma D.3.** \(^{43}\) Let \( M_\phi(\langle x_n \rangle) \) be the \( \langle \lambda_n \rangle_1^N \) weighted \( \phi \)-mean and \( M_\psi(\langle x_n \rangle) \) the \( \langle \lambda_n \rangle_1^N \) weighted \( \psi \)-mean of a tuple \( \langle x_n \rangle_1^N \).
\[
\begin{align*}
\phi \psi^{-1} \text{ is convex and } \phi \text{ is increasing } & \implies M_\phi(\langle x_n \rangle) \geq M_\psi(\langle x_n \rangle) \\
\phi \psi^{-1} \text{ is convex and } \phi \text{ is decreasing } & \implies M_\phi(\langle x_n \rangle) \leq M_\psi(\langle x_n \rangle) \\
\phi \psi^{-1} \text{ is concave and } \phi \text{ is increasing } & \implies M_\phi(\langle x_n \rangle) \leq M_\psi(\langle x_n \rangle) \\
\phi \psi^{-1} \text{ is concave and } \phi \text{ is decreasing } & \implies M_\phi(\langle x_n \rangle) \geq M_\psi(\langle x_n \rangle)
\end{align*}
\]

One of the most well known inequalities in mathematics is \textit{Minkowski’s Inequality}. In 1946, H.P. Mulholland submitted a result that generalizes Minkowski’s Inequality to an equal weighted \( \phi \)-

---


\(^{42}\) Bollobás (1999) page 5

\(^{43}\) Pečarić et al. (1992) page 107, Bollobás (1999) page 5, Hardy et al. (1952) page 75

\[\text{version 0.50e}\]

2016 Jun 16 6:40AM UTC  Daniel J. Greenhoe
mean.\(^{44}\) And Milovanović and Milovanović (1979) generalized this even further to a \textit{weighted} \(\phi\)-mean (next).

**Theorem D.13.** \(^{45}\) Let \(\phi\) be a function in \(\mathbb{R}^\mathbb{R}\).

\[
\begin{aligned}
1. \phi & \text{ is CONVEX and } \\
2. \phi & \text{ is STRICTLY MONOTONE and } \\
3. \phi(0) = 0 & \text{ and } \\
4. \log \phi \circ \exp & \text{ is CONVEX}
\end{aligned}
\]

\[
\implies \{ \phi^{-1}(\sum_{n=1}^{N} \lambda_n \phi(x_n + y_n)) \leq \phi^{-1}(\sum_{n=1}^{N} \lambda_n \phi(x_n)) + \phi^{-1}(\sum_{n=1}^{N} \lambda_n \phi(y_n)) \}
\]

**Definition D.15.** \(^{46}\) Let \(M_{\phi(x,p)}(\langle x_n \rangle)\) be the \(\langle \lambda_n \rangle^N\) \textit{weighted} \(\phi\)-mean of a NON-NEGATIVE tuple \(\langle x_n \rangle^N\). A mean \(M_{\phi(x,p)}(\langle x_n \rangle)\) is a \textbf{power mean} with parameter \(p\) if \(\phi(x) \triangleq x^p\). That is,

\[
M_{\phi(x,p)}(\langle x_n \rangle) = \left(\sum_{n=1}^{N} \lambda_n (x_n)\right)^{1/p}
\]

**Theorem D.14.** \(^{47}\) Let \(M_{\phi(x,p)}(\langle x_n \rangle)\) be the \textbf{power mean} with parameter \(p\) of an \(N\)-tuple \(\langle x_n \rangle^N\) in which the elements are NOT all equal.

\[
\begin{aligned}
M_{\phi(x,p)}(\langle x_n \rangle) & \triangleq \left(\sum_{n=1}^{N} \lambda_n (x_n)^p\right)^{1/p} & \text{is \textit{continuous} and \textit{strictly monotone} in} \ \mathbb{R}^*.
\end{aligned}
\]

\[
M_{\phi(x,p)}(\langle x_n \rangle) = \begin{cases} 
\max_{n=1,2,\ldots,N} \langle x_n \rangle & \text{for } p = +\infty \\
\prod_{n=1}^{N} x_n^{\lambda_n} & \text{for } p = 0 \\
\min_{n=1,2,\ldots,N} \langle x_n \rangle & \text{for } p = -\infty
\end{cases}
\]

\(\diamondsuit\) \textbf{Proof:}

1. Proof that \(M_{\phi(x,p)}\) is \textit{strictly monotone} in \(p\):

   (a) Let \(p\) and \(s\) be such that \(-\infty < p < s < \infty\).

   (b) Let \(\phi_p \triangleq x^p\) and \(\phi_s \triangleq x^s\). Then \(\phi_p \phi_s^{-1} = x^{\xi}\).

   (c) The composite function \(\phi_p \phi_s^{-1}\) is \textit{convex} or \textit{concave} depending on the values of \(p\) and \(s\):

   \[
   \begin{array}{c|c|c}
   p < 0 (\phi_p \text{ decreasing}) & p > 0 (\phi_p \text{ increasing}) \\
   s < 0 & \text{convex} & \text{(not possible)} \\
   s > 0 & \text{convex} & \text{concave}
   \end{array}
   \]

   (d) Therefore by Lemma D.3 (page 163),

   \[-\infty < p < s < \infty \implies M_{\phi(x,p)}(\langle x_n \rangle) < M_{\phi(x,s)}(\langle x_n \rangle)\).

2. Proof that \(M_{\phi(x,p)}\) is \textit{continuous} in \(p\) for \(p \in \mathbb{R} \setminus 0\): The sum of continuous functions is continuous. For the cases of \(p \in \{-\infty, 0, \infty\}\), see the items that follow.

\(^{44}\) Minkowski (1910) page 115, \(\bowtie\) Mulholland (1950), \(\bowtie\) Hardy et al. (1952) (Theorem 24), \(\bowtie\) Tolsted (1964) page 7, \(\bowtie\) Maligranda (1995) page 258, \(\bowtie\) Carothers (2000), page 44, \(\bowtie\) Bullen (2003) page 179

\(^{45}\) \(\bowtie\) Milovanović and Milovanović (1979), \(\bowtie\) Bullen (2003) page 306 (Theorem 9)

\(^{46}\) \(\bowtie\) Bullen (2003) page 175, \(\bowtie\) Bollobás (1999) page 6

\(^{47}\) \(\bowtie\) Bullen (2003) pages 175–177 (see also page 203), \(\bowtie\) Bollobás (1999) pages 6–8, \(\bowtie\) Bullen (1990) page 250, \(\bowtie\) Besso (1879), \(\bowtie\) Bienaymé (1840) page 68, \(\bowtie\) Brenner (1985) page 160
3. Lemma: $M_{\phi(x:p)}(\langle x_n \rangle) = \{M_{\phi(x:p)}(\langle x_n^{-1} \rangle)\}^{-1}$. Proof:

$$\{M_{\phi(x:p)}(\langle x_n^{-1} \rangle)\}^{-1} = \left\{\left(\sum_{n=1}^{N} \lambda_n (x_n^{-1})^p\right)^{\frac{1}{p}}\right\}^{-1}$$

by definition of $M_{\phi}$

$$= \left(\sum_{n=1}^{N} \lambda_n (x_n^{-1})^{-p}\right)^{\frac{1}{p}}$$

$$= M_{\phi(x:-p)}(\langle x_n \rangle)$$

by definition of $M_{\phi}$

4. Proof that $\lim_{p \to \infty} M_{\phi}(\langle x_n \rangle) = \max_{n \in Z} \{x_n\}$:

(a) Let $x_m = \max_{n \in Z} \{x_n\}$

(b) Note that $\lim_{p \to \infty} M_{\phi} \leq \max_{n \in Z} \{x_n\}$ because

$$\lim_{p \to \infty} M_{\phi}(\langle x_n \rangle) = \lim_{p \to \infty} \left(\sum_{n=1}^{N} \lambda_n x_n^p\right)^{\frac{1}{p}}$$

by definition of $M_{\phi}$

$$\leq \lim_{p \to \infty} \left(\sum_{n=1}^{N} \lambda_n x_m^p\right)^{\frac{1}{p}}$$

by definition of $x_m$ in item (4a) and because $\phi(x) \triangleq x^p$ and $\phi^{-1}$ are both increasing or both decreasing

$$= \lim_{p \to \infty} \left(x_m^p \sum_{n=1}^{N} \lambda_n\right)^{\frac{1}{p}}$$

because $x_m$ is a constant

$$= \lim_{p \to \infty} \left(x_m^p \cdot 1\right)^{\frac{1}{p}}$$

$$= x_m$$

$$= \max_{n \in Z} \{x_n\}$$

by definition of $x_m$ in item (4a)

(c) But also note that $\lim_{p \to \infty} M_{\phi} \geq \max_{n \in Z} \{x_n\}$ because

$$\lim_{p \to \infty} M_{\phi}(\langle x_n \rangle) = \lim_{p \to \infty} \left(\sum_{n=1}^{N} \lambda_n x_n^p\right)^{\frac{1}{p}}$$

by definition of $M_{\phi}$

$$\geq \lim_{p \to \infty} \left(w_m x_m^p\right)^{\frac{1}{p}}$$

by definition of $x_m$ in item (4a) and because $\phi(x) \triangleq x^p$ and $\phi^{-1}$ are both increasing or both decreasing

$$= \lim_{p \to \infty} w_m^{\frac{1}{p}} x_m^{\frac{p}{p}}$$

$$= x_m$$

$$= \max_{n \in Z} \{x_n\}$$

by definition of $x_m$ in item (4a)

(d) Combining items (b) and (c) we have $\lim_{p \to \infty} M_{\phi} = \max_{n \in Z} \{x_n\}.$
5. Proof that \( \lim_{p \to -\infty} M_{\phi}(\langle x_n \rangle) = \min_{n \in \mathbb{Z}} \langle x_n \rangle \):

\[
\lim_{p \to -\infty} M_{\phi(x,p)}(\langle x_n \rangle) = \lim_{p \to -\infty} M_{\phi(x,-p)}(\langle x_n \rangle) = \lim_{p \to -\infty} \left( M_{\phi(x,p)}(\langle x_{n-1} \rangle) \right)^{-1} = \frac{1}{\lim_{p \to -\infty} M_{\phi(x,p)}(\langle x_{n-1} \rangle)} = \frac{1}{\lim_{p \to -\infty} M_{\phi(x,p)}(\langle x_{n-1} \rangle) - 1} = \frac{1}{\max_{n \in \mathbb{Z}} \langle x_{n-1} \rangle - 1} = \min_{n \in \mathbb{Z}} \langle x_n \rangle
data by change of variable \( p \)

\[ \text{by Lemma in item (3) page 165} \]

\[ \text{by property of \( \lim \) 48} \]

\[ \text{by item (4)} \]

6. Proof that \( \lim_{p \to 0} M_{\phi}(\langle x_n \rangle) = \prod_{n=1}^{N} x_n^{\lambda_n} \):

\[
\lim_{p \to 0} M_{\phi}(\langle x_n \rangle) = \lim_{p \to 0} \exp \left\{ \ln \left\{ M_{\phi}(\langle x_n \rangle) \right\} \right\}
\]

\[ = \lim_{p \to 0} \exp \left\{ \ln \left( \left( \sum_{n=1}^{N} \lambda_n (x_n^p) \right)^{\frac{1}{p}} \right) \right\} \]

\[ = \exp \left\{ \frac{\frac{\partial}{\partial p} \ln \left( \sum_{n=1}^{N} \lambda_n (x_n^p) \right)}{\frac{\partial}{\partial p} p} \right\}_{p=0} \]

\[ = \exp \left\{ \frac{\sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial p} (x_n^p)}{\sum_{n=1}^{N} \lambda_n (x_n^p)} \right\}_{p=0}
\]

\[ = \exp \left\{ \frac{\sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial p} (\ln (x_n))}{1} \right\}_{p=0}
\]

\[ = \exp \left\{ \sum_{n=1}^{N} \lambda_n \frac{\partial}{\partial p} (\ln (x_n)) \right\}_{p=0} = \exp \left\{ \sum_{n=1}^{N} \lambda_n \ln (x_n) \right\}_{p=0}
\]

\[ = \exp \left\{ \sum_{n=1}^{N} \lambda_n \exp (\ln (x_n)) \right\}_{p=0} = \exp \left\{ \sum_{n=1}^{N} \lambda_n (x_n) \right\}_{p=0}
\]

\[ = \exp \left\{ \sum_{n=1}^{N} \lambda_n \exp (\ln (x_n)) \right\}_{p=0} = \exp \left\{ \sum_{n=1}^{N} \lambda_n \ln (x_n) \right\}_{p=0}
\]

\[ = \exp \left\{ \prod_{n=1}^{N} x_n^{\lambda_n} \right\}_{p=0} = \prod_{n=1}^{N} x_n^{\lambda_n}
\]

\[ \text{by definition of \( M_{\phi} \)} \]

\[ \text{by l'Hôpital's rule 49} \]

\[ \text{by \( \lim \) 48} \]

\[ \text{by Rudin (1976) page 85 (4.4 Theorem)} \]
**Corollary D.2.** \(50\) Let \(\langle x_n \rangle_1^N \) be a tuple. Let \(\langle \lambda_n \rangle_1^N \) be a tuple of weighting values such that \(\sum_{n=1}^N \lambda_n = 1\). 

\[
\min \|x_n\| \leq \left( \sum_{n=1}^N \frac{\lambda_n}{x_n} \right)^{-1} \leq \prod_{n=1}^N x_n^{\lambda_n} \leq \sum_{n=1}^N \lambda_n x_n \leq \max \|x_n\|
\]

\(\text{harmonic mean} \quad \text{geometric mean} \quad \text{arithmetic mean}\)

\(\text{Proof:}\)

1. These five means are all special cases of the power mean \(M_{\phi(x; p)}\) (Definition D.15 page 164)
   - \(p = \infty\): \(\max \|x_n\|\)
   - \(p = 1\): arithmetic mean
   - \(p = 0\): geometric mean
   - \(p = -1\): harmonic mean
   - \(p = -\infty\): \(\min \|x_n\|\)

2. The inequalities follow directly from Theorem D.14 (page 164).

3. Generalized AM-GM inequality: If one is only concerned with the arithmetic mean and geometric mean, their relationship can be established directly using **Jensen’s Inequality**:

\[
\sum_{n=1}^N \lambda_n x_n = b^{\log_b \left( \sum_{n=1}^N \lambda_n x_n \right)} \\
\geq b^{\left( \sum_{n=1}^N \lambda_n \log_b x_n \right)} \quad \text{by Jensen’s Inequality} \quad \text{(Theorem D.12 page 163)} \\
= \prod_{n=1}^N b^{\left( \lambda_n \log_b x_n \right)} = \prod_{n=1}^N b^{\left( \log_b x_n \right) \lambda_n} = \prod_{n=1}^N x_n^{\lambda_n}
\]

\(\text{D.3.4 Inequalities}\)

**Lemma D.4** (Young’s Inequality). \(51\)

\[
xy < \frac{x^p}{p} + \frac{y^q}{q} \quad \text{with} \quad \frac{1}{p} + \frac{1}{q} = 1 \quad \forall 1 < p < \infty, \quad x, y \geq 0, \quad \text{but} \quad y \neq x^{p-1}
\]

\[
xy = \frac{x^p}{p} + \frac{y^q}{q} \quad \text{with} \quad \frac{1}{p} + \frac{1}{q} = 1 \quad \forall 1 < p < \infty, \quad x, y \geq 0, \quad \text{and} \quad y = x^{p-1}
\]

**Theorem D.15** (Minkowski’s Inequality for sequences). \(52\) Let \(\langle x_n \in \mathbb{C} \rangle_1^N \) and \(\langle y_n \in \mathbb{C} \rangle_1^N \) be complex \(N\)-tuples.

\[
\left( \sum_{n=1}^N |x_n + y_n|^p \right)^{\frac{1}{p}} \leq \left( \sum_{n=1}^N |x_n|^p \right)^{\frac{1}{p}} + \left( \sum_{n=1}^N |y_n|^p \right)^{\frac{1}{p}} \quad \forall 1 < p < \infty
\]


\(51\) **Young (1912) page 226, **Hardy et al. (1952) (Theorem 24), **Tolsted (1964) page 5, **Maligranda (1995) page 257, **Carothers (2000), page 43

\(52\) **Minkowski (1910), page 115, **Hardy et al. (1952) (Theorem 24), **Maligranda (1995) page 258, **Tolsted (1964), page 7, **Carothers (2000), page 44, **Bullen (2003) page 179
D.4 Topological Spaces

“Nevertheless I should not pass over in silence the fact that today the feeling among mathematicians is beginning to spread that the fertility of these abstracting methods is approaching exhaustion. The case is this: that all these nice general concepts do not fall into our laps by themselves. But definite concrete problems were first conquered in their undivided complexity, singlehanded by brute force, so to speak. Only afterwards the axiomaticians came along and stated: Instead of breaking the door with all your might and bruising your hands, you should have constructed such and such a key of skill, and by it you would have been able to open the door quite smoothly. But they can construct the key only because they are able, after the breaking in was successful, to study the lock from within and without. Before you can generalize, formalize, and axiomatize, there must be a mathematical substance.”

Hermann Weyl (1885–1955); mathematician, theoretical physicist, and philosopher

Definition D.16. 54 Let \( \Gamma \) be a set with an arbitrary (possibly uncountable) number of elements. Let \( 2^X \) be the power set of a set \( X \) (Definition 1.8 page 6). A family of sets \( T \subseteq 2^X \) is a topology on \( X \) if

1. \( \emptyset, X \in T \)
2. \( X \in T \)
3. \( U, V \in T \implies U \cap V \in T \)
4. \( \{ U_\gamma \mid \gamma \in \Gamma \} \subseteq T \implies \bigcup_{\gamma \in \Gamma} U_\gamma \in T \).

The ordered pair \((X, T)\) is a topological space if \( T \) is a topology on \( X \). A set \( U \) is open in \((X, T)\) if \( U \) is any element of \( T \). A set \( D \) is closed in \((X, T)\) if \( D^c \) is open in \((X, T)\).

Just as the power set \( 2^X \) and the set \( \{ \emptyset, X \} \) are algebras of sets on a set \( X \), so also are these sets topologies on \( X \) (next example):

Example D.12. 55 Let \( \mathcal{T}(X) \) be the set of topologies on a set \( X \) and \( 2^X \) the power set (Definition 1.8 page 6) on \( X \).

- \( \{ \emptyset, X \} \) is a topology in \( \mathcal{T}(X) \) (indiscrete topology or trivial topology)
- \( 2^X \) is a topology in \( \mathcal{T}(X) \) (discrete topology)

Definition D.17. 56 Let \((X, T)\) be a topological space. A set \( B \subseteq 2^X \) is a base for \( T \) if

1. \( B \subseteq T \)
2. \( \forall U \in T, \exists \{ B_\gamma \in B \} \text{ such that } U = \bigcup_\gamma B_\gamma \)

Theorem D.16. 57 Let \((X, T)\) be a topological space. Let \( B \subseteq 2^X \) such that \( B \subseteq 2^X \).

\[
\{ B \text{ is a base for } T \} \iff \left\{ \begin{array}{l}
\text{For every } x \in X \text{ and every open set } U \text{ containing } x,
\text{ there exists } B_x \in B \text{ such that } x \in B_x \subseteq U.
\end{array} \right\}
\]

Theorem D.17. 58 Let \((X, T)\) be a topological space (Definition D.16 page 168) and \( B \subseteq 2^X \).

\[
\text{\emph{B is a base for } (X, T) \iff \begin{cases} 
1. x \in X \quad \Rightarrow \exists B_x \in B \text{ such that } x \in B_x \subseteq U \\
2. B_1, B_2 \in B \quad \Rightarrow B_1 \cap B_2 \in B
\end{cases}}
\]

53 quote: 
Weyl (1935) page 14 (H. Weyl, quoting himself from “a conference on topology and abstract algebra as two ways of mathematical understanding, in 1931”), image: https://en.wikipedia.org/wiki/File:Hermann_Weyl_ETH-Bib_Portr_00890.jpg: “This work is free and may be used by anyone for any purpose.”

54 Munkres (2000) page 76, Riesz (1909), Hausdorff (1914), Tietze (1923), Hausdorff (1937) page 258


56 Joshi (1983) page 92 (3.1 Definition), Davis (2005) page 46 (Definition 4.15)

57 Joshi (1983) pages 92–93 (3.2 Proposition), Davis (2005) page 46

58 Bollobás (1999) page 19
Example D.13. \(^{59}\) Let \((X, d)\) be a metric space. The set \(B \triangleq \{ B(x, r) \mid x \in X, \ r \in \mathbb{N} \}\) (the set of all open balls in \((X, d)\)) is a base for a topology on \((X, d)\).

Example D.14 (the standard topology on the real line). \(^{60}\) The set \(B \triangleq \{(a : b) \mid a, b \in \mathbb{R}, \ a < b \}\) is a base for the metric space \((\mathbb{R}, \ |b - a|)\) (the usual metric space on \(\mathbb{R}\)).

Definition D.18. \(^{61}\) Let \((X, T)\) be a topological space (Definition D.16 page 168). Let \(2^X\) be the power set of \(X\).

The set \(A^-\) is the closure of \(A \subseteq 2^X\) if \(A^- \triangleq \bigcap \{ D \subseteq 2^X \mid A \subseteq D \text{ and } D \text{ is closed} \}\).

The set \(A^+\) is the interior of \(A \subseteq 2^X\) if \(A^+ \triangleq \bigcup \{ U \subseteq 2^X \mid U \subseteq A \text{ and } U \text{ is open} \}\).

A point \(x\) is a closure point of \(A\) if \(x \in A^-\).

A point \(x\) is an interior point of \(A\) if \(x \in A^+\).

A point \(x\) is an accumulation point of \(A\) if \(x \in (A \setminus \{ x \})^-\).

A point \(x\) in \(A^-\) is a point of adherence in \(A\) or is adherent to \(A\) if \(x \in A^-\).

Proposition D.2. \(^{62}\) Let \((X, T)\) be a topological space (Definition D.16 page 168). Let \(A^-\) be the closure, \(A^+\) the interior, and \(\partial A\) the boundary of a set \(A\). Let \(2^X\) be the power set of \(X\).

1. \(A^-\) is closed \(\forall A \subseteq 2^X\).
2. \(A^+\) is open \(\forall A \subseteq 2^X\).

Lemma D.5. \(^{63}\) Let \(A^-\) be the closure, \(A^+\) the interior, and \(\partial A\) the boundary of a set \(A\) in a topological space \((X, T)\). Let \(2^X\) be the power set of \(X\).

1. \(A^+ \subseteq A \subseteq A^-\) \(\forall A \subseteq 2^X\).
2. \(A = A^+ \iff A\) is open \(\forall A \subseteq 2^X\).
3. \(A = A^- \iff A\) is closed \(\forall A \subseteq 2^X\).

Definition D.19. \(^{64}\) Let \((X, T_X)\) and \((Y, T_Y)\) be topological spaces (Definition D.16 page 168). Let \(f\) be a function in \(Y^X\). A function \(f \in Y^X\) is continuous if \(\forall U \subseteq T_Y\) \(\text{open in } (Y, T_Y)\), \(f^{-1}(U) \subseteq T_X\) \(\text{open in } (X, T_X)\).

A function is discontinuous in \((X, T_X)\) if it is not continuous in \((X, T_X)\).

![Figure D.3: continuous/discontinuous functions](Example D.15 page 169)

Example D.15. Some continuous/discontinuous functions are illustrated in Figure D.3 (page 169).

Definition D.19 (previous definition) defines continuity using open sets. Continuity can alternatively be defined using closed sets or closure (next theorem).

---

\(^{59}\) Davis (2005) page 46 (Example 4.16)

\(^{60}\) Munkres (2000) page 81, Davis (2005) page 46 (Example 4.16)


\(^{62}\) McCarty (1967) page 90 (IV.1 theorem)

\(^{63}\) McCarty (1967) pages 90–91 (IV.1 theorem), Aliprantis and Burkinshaw (1998) page 59

\(^{64}\) Davis (2005) page 34
Theorem D.18. 65 Let \((X, T)\) and \((Y, S)\) be topological spaces. Let \(f\) be a function in \(Y^X\).

The following are equivalent:

1. \(f\) is CONTINUOUS
2. \(B\) is closed in \((Y, S)\) \(\implies\) \(f^{-1}(B)\) is closed in \((X, T)\) \(\forall B \subseteq Y\)
3. \(f(A^-) \subseteq f(A^-)\) \(\forall A \subseteq X\)
4. \(f^{-1}(B^-) \subseteq f^{-1}(B^-)\) \(\forall B \subseteq Y\)

Remark D.1. A word of warning about defining continuity in terms of topological spaces—continuity is defined in terms of a pair of topological spaces, and whether function is continuous or discontinuous in general depends very heavily on the selection of these spaces. This is illustrated in Proposition D.3 (next). The ramification of this is that when declaring a function to be continuous or discontinuous, one must make clear the assumed topological spaces.

Proposition D.3. 66 Let \((X, T)\) and \((Y, S)\) be topological spaces. Let \(f\) be a function in \((Y, S)^{(X,T)}\).

1. \(T\) is the DISCRETE TOPOLOGY \(\implies\) \(f\) is CONTINUOUS \(\forall f \in (Y,S)^{(X,T)}\)
2. \(S\) is the INDISCRETE TOPOLOGY \(\implies\) \(f\) is CONTINUOUS \(\forall f \in (Y,S)^{(X,T)}\)

Definition D.20. 67 Let \((X, T)\) be a TOPOLOGICAL SPACE (Definition D.16 page 168). A sequence \((x_n)_{n \in \mathbb{Z}}\) converges in \((X, T)\) to a point \(x\) if for each open set (Definition D.16 page 168) \(U \in T\) that contains \(x\) there exists \(N \in \mathbb{N}\) such that \(x_n \in U\) for all \(n > N\).

This condition can be expressed in any of the following forms:

1. The limit of the sequence \((x_n)\) is \(x\).
2. The sequence \((x_n)\) is convergent with limit \(x\).
3. \(\lim_{n \to \infty} (x_n) = x\).
4. \((x_n) \to x\).

A sequence that converges is convergent. A sequence that does not converge is said to diverge, or is divergent. An element \(x \in A\) is a limit point of \(A\) if it is the limit of some \(A\)-valued sequence \((x_n)_{n \in A}\).

Example D.16. 68 Let \((X, T_{31})\) be a topological space where \(X \triangleq \{x, y, z\}\) and

\[ T_{31} \triangleq \{\emptyset, \{x\}, \{x, y\}, \{x, z\}, \{x, y, z\}\}. \]

In this space, the sequence \((x, x, x, \ldots)\) converges to \(x\). But this sequence also converges to both \(y\) and \(z\) because \(x\) is in every open set (Definition D.16 page 168) that contains \(y\) and \(x\) is in every open set that contains \(z\). So, the limit (Definition D.20 page 170) of the sequence is not unique.

Example D.17. In contrast to the low resolution topological space of Example D.16, the limit of the sequence \((x, x, x, \ldots)\) is unique in a topological space with sufficiently high resolution with respect to \(y\) and \(z\) such as the following: Define a topological space \((X, T_{56})\) where \(X \triangleq \{x, y, z\}\) and

\[ T_{56} \triangleq \{\emptyset, \{y\}, \{z\}, \{x, y\}, \{y, z\}, \{x, y, z\}\}. \]

In this space, the sequence \((x, x, x, \ldots)\) converges to \(x\) only. The sequence does not converge to \(y\) or \(z\) because there are open sets (Definition D.16 page 168) containing \(y\) or \(z\) that do not contain \(x\) (the open sets \{\(y\}\}, \{\(z\}\}, and \(\{x, y, z\}\)).

Theorem D.19 (The Closed Set Theorem). 69 Let \((X, T)\) be a topological space. Let \(A\) be a subset of \(X\) (\(A \subseteq X\)). Let \(A^-\) be the CLOSURE (Definition D.18 page 169) of \(A\) in \((X, T)\).

\[ A \text{ is closed in } (X, T) \iff \begin{cases} \text{Every } A\text{-valued sequence } (x_n \in A)_{n \in \mathbb{Z}} \\ \text{that converges in } (X, T) \text{ has its limit in } A \end{cases} \]


66 Crossley (2006) page 18 (Proposition 3.9), Ponnusamy (2002) page 98 (2.64. Theorem.)

67 Joshi (1983) page 83 ((3.1) Definition), Leathem (1905), page 13 ("\(\rightarrow\)" symbol, section III.11)

68 Munkres (2000) page 98 (Hausdorff Spaces)

**Theorem D.20.** Let $(X, T)$ and $(Y, S)$ be a topological spaces. Let $f$ be a function in $(Y, S)^{(X, T)}$.

\[
\begin{align*}
\begin{cases}
\text{$f$ is continuous in $(Y, S)^{(X, T)}$} \\
(\text{Definition D.19 page 169})
\end{cases} & \iff \\
\begin{cases}
\{x_n\} \to x \implies f((x_n)) \to f(x) \\
(\text{Definition D.20 page 170})
\end{cases}
\end{align*}
\]

\text{Inverse image characterization of continuity} \iff \text{Sequential characterization of continuity}

\text{Proof:}

1. Proof for the $\implies$ case (proof by contradiction):

   (a) Let $U$ be an open set in $(Y, T)$ that contains $f(x)$ but for which there exists no $N$ such that $f(x_n) \in U$ for all $n > N$.

   (b) Note that the set $f^{-1}(U)$ is also open by the continuity hypothesis.

   (c) If $\{x_n\} \to x$, then

   \[
   f((x_n)) \to f(x) \implies \text{there exists } N \text{ such that } f(x_n) \in U \text{ for all } n > N \quad \text{by Definition D.20}
   \]

   \[
   \implies \text{there exists } M \text{ such that } x_n \in f^{-1}(U) \text{ for all } n > M \quad \text{by definition of } f^{-1}
   \]

   \[
   \implies \{x_n\} \to x \quad \text{by continuity hyp. and def. of convergence (Definition D.20 page 170)}
   \]

   \[
   \implies \text{contradiction of } \{x_n\} \to x \text{ hypothesis}
   \]

   \[
   \implies f((x_n)) \to f(x)
   \]

2. Proof for the $\iff$ case (proof by contradiction):

   (a) Let $D$ be a closed set in $(Y, S)$.

   (b) Suppose $f^{-1}(D)$ is not closed…

   (c) then by the closed set theorem (Theorem D.19 page 170), there must exist a convergent sequence $\{x_n\}$ in $(X, T)$, but with limit $x$ not in $f^{-1}(D)$.

   (d) Note that $f(x)$ must be in $D$. Proof:

      i. by definition of $D$ and $f$, $f((x_n))$ is in $D$

      ii. by left hypothesis, the sequence $f((x_n))$ is convergent with limit $f(x)$

      iii. by closed set theorem (Theorem D.19 page 170), $f(x)$ must be in $D$.

   (e) Because $f(x) \in D$, it must be true that $x \in f^{-1}(D)$.

   (f) But this is a contradiction to item (2c) (page 171), and so item (2b) (page 171) must be wrong, and $f^{-1}(D)$ must be closed.

   (g) And so by Theorem D.18 (page 170), $f$ is continuous.

---

70 Ponnusamy (2002) pages 94–96 ("2.59. Proposition."; in the context of metric spaces; includes the "inverse image characterization of continuity" and "sequential characterization of continuity" terminology; this terminology does not seem to be widely used in the literature in general, but has been adopted for use in this text)
E.1 Lagrange interpolation

Definition E.1. The Lagrange polynomial \( L_{P,n}(x) \) with respect to the \( n + 1 \) points \( P = \{(x_k, y_k) | k = 0, 1, 2, \ldots, n\} \) is defined as

\[
L_{P,n}(x) = \sum_{k=0}^{n} y_k \prod_{m \neq k} \frac{x - x_m}{x_k - x_m}
\]

Proposition E.1. Let \( L_{P,n}(x) \) be the Lagrange polynomial with respect to the points \( P = \{(x_k, y_k) | k = 0, 1, 2, \ldots, n\} \).
1. \( L_{P,n}(x) \) is an \( n \)th order polynomial.
2. \( L_{P,n}(x) \) intersects all \( n + 1 \) points in \( P \).

Example E.1 (Lagrange interpolation). The Lagrange polynomial \( L_{P,3}(x) \) with respect to the 4 points \( P = \{(-2, 1), (-1, 3), (3, 2), (5, 4)\} \) is

\[
L_{P,3}(x) = \frac{79}{840} x^3 + \frac{-378}{840} x^2 + \frac{-7}{840} x + \frac{2970}{840}
\]
E.2 Newton Interpolation

Definition E.2. 3 The Newton polynomial \( N_{P,n}(x) \) with respect to the \( n+1 \) points 
P = \{ (x_k, y_k) | k = 0, 1, 2, \ldots, n \} is defined as 
\[
N_{P,n}(x) = \sum_{k=0}^{n} a_k \prod_{m=0}^{n-k} (x - x_m)
\]

Proposition E.2. Let \( N_{P,n}(x) \) be the Newton polynomial with respect to the points 
P = \{ (x_k, y_k) | k = 0, 1, 2, \ldots, n \}.

1. \( N_{P,n}(x) \) is an \( n \)th order polynomial.

2. \( N_{P,n}(x) \) intersects all \( n + 1 \) points in \( P \).

Example E.2 (Newton polynomial interpolation). The Newton polynomial \( N_{P,3}(x) \) with respect to the 4 points \( P = \{ (-2, 1), (-1, 3), (3, 2), (5, 4) \} \) is 
\[
N_{P,3}(x) = \frac{79}{840} x^3 + \frac{337}{280} x^2 + \frac{7}{40} x + \frac{3970}{41760}
\]

Theorem:

\[ N_{P,3}(x) = \sum_{k=0}^{n} a_k \prod_{m=1}^{k} (x - x_m) \]

\[ = a_0 + a_1(x - x_0) + a_2(x - x_0)(x - x_1) + a_3(x - x_0)(x - x_1)(x - x_2) \]

\[ = a_0 + a_1(x + 2) + a_2(x + 2)(x + 1) + a_3(x + 2)(x + 1)(x + 3) \]

\[ = a_0 + a_1(x + 2) + a_2(x^2 + 3x + 2) + a_3(x^3 - 7x - 6) \]

\[ = x^3(a_3) + x^2(a_2) + x(-7a_3 + 3a_2 + a_1) + (-6a_3 + 2a_2 + 2a_1 + a_0) \]

\[ = \begin{bmatrix}
  a_0 & a_1 & a_2 & a_3
\end{bmatrix}
\begin{bmatrix}
  1 & 0 & 0 & 0 \\
  2 & 1 & 0 & 0 \\
  2 & 3 & 1 & 0 \\
  -6 & -7 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
  x \\
  x^2 \\
  x^3
\end{bmatrix} \]

\[ \begin{bmatrix}
  1 & 3 & 2 & 4
\end{bmatrix} = \begin{bmatrix}
  y_0 \\
  y_1 \\
  y_2 \\
  y_3
\end{bmatrix} \]

\[ = \begin{bmatrix}
  1 & 0 & 0 & 0 & 0 \\
  1 & (x_1 - x_0) & 0 & 0 \\
  1 & (x_2 - x_0)(x_2 - x_1) & 0 \\
  1 & (x_3 - x_0)(x_3 - x_1)(x_3 - x_2) & 0
\end{bmatrix}
\begin{bmatrix}
  a_0 \\
  a_1 \\
  a_2 \\
  a_3
\end{bmatrix} \]

\[ \begin{bmatrix}
  1 & 0 & 0 & 0 \\
  1 & 0 & 0 & 0 \\
  1 & 1 & 0 & 0 \\
  1 & 5 & 20 & 0
\end{bmatrix} = \begin{bmatrix}
  a_0 \\
  a_1 \\
  a_2 \\
  a_3
\end{bmatrix} \]

\[ \begin{bmatrix}
  1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
  1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
  1 & 5 & 20 & 0 & 0 & 0 & 1 & 0 \\
  1 & 7 & 42 & 84 & 0 & 0 & 0 & 1
\end{bmatrix} = \begin{bmatrix}
  1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \\
  0 & 5 & 20 & 0 & -1 & 0 & 1 & 0 \\
  0 & 7 & 42 & 84 & -1 & 0 & 0 & 1
\end{bmatrix} \]

\[ = \begin{bmatrix}
  1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \\
  0 & 0 & 20 & 0 & 4 & -5 & 1 & 0 \\
  0 & 0 & 42 & 84 & 6 & -7 & 0 & 1
\end{bmatrix} \]

\[ = \begin{bmatrix}
  1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \\
  0 & 0 & 1 & 0 & \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} & 0 \\
  0 & 0 & 42 & 84 & 6 & -7 & 0 & 1
\end{bmatrix} \]

\[ = \begin{bmatrix}
  1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \\
  0 & 0 & 1 & 0 & \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} & 0 \\
  0 & 0 & 84 & -\frac{12}{3} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 1
\end{bmatrix} \]

\[ = \begin{bmatrix}
  1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 & -1 & 1 & 0 & 0 \\
  0 & 0 & 1 & 0 & \frac{1}{3} & -\frac{1}{3} & \frac{1}{3} & 0 \\
  0 & 0 & 84 & -\frac{12}{3} & -\frac{1}{4} & \frac{1}{4} & \frac{1}{4} & 1
\end{bmatrix} \]
\[
\begin{bmatrix}
\alpha_0 \\
\alpha_1 \\
\alpha_2 \\
\alpha_3
\end{bmatrix}
= \begin{bmatrix}
1 & 0 & 0 & 0 \\
-1 & 1 & 0 & 0 \\
\frac{4}{24} & -\frac{5}{35} & \frac{1}{10} & 0 \\
0 & 0 & 1 & -\frac{4}{24} & -\frac{5}{35} & \frac{1}{10} & 0
\end{bmatrix}
\begin{bmatrix}
1 \\
3 \\
2 \\
4
\end{bmatrix}
\]

\[
N_{p,3}(x) = \begin{bmatrix}
\alpha_0 \\
\alpha_1 \\
\alpha_2 \\
\alpha_3
\end{bmatrix}
= \begin{bmatrix}
1 & 0 & 0 & 0 \\
2 & 1 & 0 & 0 \\
2 & 3 & 1 & 0 \\
-6 & -7 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
1 \\
x \\
x^2 \\
x^3
\end{bmatrix}
\]

\[
= \begin{bmatrix}
1 \\
2 \\
-\frac{9}{20} \\
\frac{79}{840}
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 & 0 \\
2 & 1 & 0 & 0 \\
2 & 3 & 1 & 0 \\
-6 & -7 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
1 \\
x \\
x^2 \\
x^3
\end{bmatrix}
\]

\[
= \begin{bmatrix}
1 \\
1 + 4 & -\frac{9}{10} & -\frac{79}{840} \\
2 & -\frac{27}{20} & -\frac{79}{140} & \frac{9}{20} \\
-6 & -7 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
1 \\
x \\
x^2 \\
x^3
\end{bmatrix}
\]

\[
= \frac{79}{840}x^3 - \frac{378}{840}x^2 - \frac{7}{840}x + \frac{2970}{840}
\]
APPENDIX F

C++ SOURCE CODE SUPPORT

This document seeks to conform to the principles of *Reproducible Research* as detailed at http://reproducibleresearch.net/.

This section contains a partial C++ source code listing for ssp.exe, written by the author of this paper, which produced the \TeX files used to generate the 128 or so data plot files presented in CHAP**TER 3. The complete and downloadable source code for ssp.exe is to accompany any online version of this document.

There are some who might hesitate to use computer simulation to demonstrate mathematical concepts. And arguably their concern is not unfounded. There is in fact evidence to suggest that while the invention of the printing press has greatly assisted the progress of mathematical discovery, the introduction of computers has harmed it. Evidence of this hypothesis is given in Figure F.1 (page 178).\(^1\) This graph shows the number of “notable” mathematicians alive during the last 3000 years; Here a “notable mathematician” is defined as one whose name appears in *Saint Andrew's University's Who Was There* website.\(^2\) Note the following:

- The number of mathematicians starts to exponentially increase at about the time of Gutenberg’s invention of the printing press— that is, when information of discoveries and results could be widely and economically circulated.\(^3\)
- There is another increase after the invention of the slide rule in the early 1600s— that is, when computational power increased.
- There are huge increases around the time of the first and second industrial revolutions— that is, when there were many *applications* that called for mathematical solutions.
- After the invention of the pocket scientific calculator in 1972 and home IBM PC in 1981— machines that could often make hard-core mathematical analysis unnecessary in real-world applications— there was a huge drop in the number of mathematicians.

### F.1 Symbolic sequence routines

```cpp
/*==============================================================================
 * Daniel J. Greenhoe
 */

\(^1\)Data for Figure F.1 (page 178) extracted from http://www-history.mcs.st-andrews.ac.uk/Timelines/WhoWasThere.html
\(^2\)http://www-history.mcs.st-andrews.ac.uk/Timelines/WhoWasThere.html
\(^3\)This point is also made by Resnikoff and Wells in Resnikoff and Wells (1984), page 9.
```
Figure F.1: Number of notable mathematicians alive over time

```
* header file for routines for char sequence functions

class symseq {
private:
long N;
char *x;
public:
    symseq(const long M); //constructor initializing to '.'
symseq(const long M, const unsigned seed, const char *symbols); //constructor initializing using seed
void clear(void);  //fill sequence with the value 'A'
char get(const long n);  //get a value from x at location n
char get(const long n, char *symbols);  //get a value from x at location n but exit if not in
    <symbols> string
void put(long n, char symbol);
void put(const long start, const long end, const char symbol);
const long getN(void) {const long M=(const long)N; return M;}  //get N
void downsample(int M, symseq *y);  //downsample by a factor of <M> and write to <y>
void list(const long start, const long end, const char *str1, const char *str2, const int int display, FILE *fptr);  
void list(const long start, const long end, FILE *fptr){list(start,end, "","",1,fptr);}  
void list(int long start, const long end, const char *str1, const char *str2, FILE *fptr){  
    list(start,end,str1,str2,1,fptr);
}
void list(const long start, long end){list(start,end,"","",NULL);}  //list contents of sequence
void list(void){  
    list(0,-1,"","",NULL);}  //list contents of sequence
void list(const long start){  
    list(start,N-1,"","",NULL);}  //list contents of sequence
void shiftL(long n);  //shift symseq n elements to the left
void shiftR(long n);  //shift symseq n elements to the right
void prngseed(unsigned seed)(rand(seed));  //
void randomize(const char *symbols);  //randomize using a list of symbols
void randomize(const unsigned seed, const char *symbols){prngseed(seed); randomize(symbols);  
void operator=(symseq y);  //x=y
void operator<==(const long n){shiftR(n);  //shift symseq n elements to the right
void operator<=(const long n){shiftL(n);  //shift symseq n elements to the left
int operator=(symseq y);  //test if x=y; 1 if yes, 0 if no
};

extern long cmp(const symseq *x, const symseq *y, int showdiff, FILE *fptr);
void copy(const long start, const long end, const symseq *x, const symseq *y);
void downsample(int M, symseq *x, symseq *y);
```
* constructor initializing symseq to '. *

symseq::symseq(long M)
{
    long n;
    void *memptr;
    N=M;
    memptr=malloc(N*sizeof(char));
    if (memptr==NULL) {
        fprintf(stderr, "symseq::symseq memory allocation for %ld elements failed\n" ,M);
        exit(EXIT_FAILURE);
    }
    x = (char *)memptr;
    for(n=0; n<N; n++)x[n]='.';
}

/* constructor initializing symseq to '. '

symseq::symseq(const long M, const unsiged seed, const char *symbols)
{
    long n;
    void *memptr;
    N=M;
    memptr=malloc(N*sizeof(char));
    if (memptr==NULL) {
        fprintf(stderr, "symseq::symseq memory allocation for %ld elements failed\n" ,M);
        exit(EXIT_FAILURE);
    }
    x = (char *)memptr;
    randomize(seed,symbols);
}

/* fill the symseq with the value '.'

void symseq::clear(void)
{
    long n;
    for(n=0; n<N; n++)x[n]='.';
}

/* get a symbol from the symseq x at location n

char symseq::get(const long n)
{
    if(n<0 || n>=N) /*domain check
        fprintf(stderr, "ERROR using symseq::get(n) : n=%ld outside the domain [0:%ld] of the
        sequence.\n" ,n,N);
        exit(EXIT_FAILURE);
    return x[n];
}

/* get a symbol from the symseq x at location n
* but exit if symbol is not in the string <range>
* example: symbol get(n,"ABCDEF");

char symseq::get(const long n,char *range)
{
    int M;
    int match;
    char *sptr;
    char symbol;
    if(n<0 || n>=N) /*domain check
        fprintf(stderr, "ERROR using symseq::get(n) : n=%ld outside the domain [0:%ld] of the
        sequence.\n" ,n,N);
        exit(EXIT_FAILURE);
    symbol = x[n];
    for(match=0,sptr=range; *sptr!='\0'; sptr++) if (symbol==*sptr) match=1;
    if (!match) /*range check
        fprintf(stderr, "\nERROR using symbol=symseq::get(%ld) : symbol='\"%c\"' (0x%02X) outside the range [%s]
        of the sequence.\n" ,n,symbol,symbol,range);
        exit(EXIT_FAILURE);
    return symbol;
}
* put a single value from the symseq at location n
  
void symseq::put(long n, char symbol){
  int rval;
  if(n<0 || n>=N) // domain check
    fprintf(stderr,"ERROR using symseq::put(n): n=%ld outside the domain [0,%ld] of the
               sequence \n",n,N-1);
    exit(EXIT_FAILURE);
  x[n]=symbol;
}

/*
   * put the symbol <symbol> into the sequence <x>
   * from location <start> to location <end>
   */
void symseq::put(const long start, const long end, const char symbol){
  long n;
  for(n=start;n<=end;n++) put(n,symbol);
}

/*
   * list contents of dieseq
   */
void symseq::list(const long start, const long end, const char *str1, const char *str2, const int display, FILE *fptr){
  long n,m;
  if(strlen(str1)!=0) {
    if(display) printf("%s",str1);
    if(fptr!=NULL) fprintf(fptr,"%s",str1);
  }
  for(n=start,m=1; n<=end; n++,m++) {
    if(display) printf("%c",get(n));
    if(m%10==0&&display) printf("\n");
    else if(m%10==0) printf("\n");
    if(fptr!=NULL) 
      fprintf(fptr,"%c",get(n));
    if(m%10==0) fprintf(fptr,"\n");
    else if(m%10==0) fprintf(fptr,"\n");
  }
}

/*
   * shift symseq n elements to the right inserting zeros on the left
   * example: if x = [ a b c d e f ] (N=6), then shiftR(2) results in
   *           x = [ 0 0 a b c d ] (N=6).
   */
void symseq::shiftR(long n){
  long m;
  for(m=N-1,m-n>=0;m-->) x[m]=x[m-n];
  for(m=0;m<n;m++) x[m]='.';
}

/*
   * shift symseq n elements to the left inserting zeros on the right
   * example: if x = [ a b c d e f ] (N=6), then shiftL(2) results in
   *           x = [ c d e f 0 0 ] (N=6).
   */
void symseq::shiftL(long n){
  long m;
  for(m=0;m<N-n;m++) x[m]=x[m+n];
  for(m=0;m<N;m++) x[m]='.';
}

/*
   * fill the sequence with uniformly distributed pseudo-random symbols
   * from the string <symbols>
   */
void symseq::randomize(const char *symbols){
  const long N=getN();
  const int M=strlen(symbols);
  int r,i;
}
long n;
for(n=0; n<N; n++) {
  r=rand();
i = r%M;
  put[n, symbols[i]];
}
*/
* operators
* operator symseq x = symseq y
*/
void symseq::operator=(symseq y) {
  const long M=y.getN();
  long n;
  if(N!=M) {
    fprintf(stderr, "ERROR using symseq x = symseq y operation: size of x (%ld) does not equal size of y (%ld)\n",N,M);
    exit(EXIT_FAILURE);
  }
  for(n=0;n<N;n++) x[n]=y.get(n);
}
*/
* operator symseq x == symseq y
* compare x and y; return 1 if the same, 0 if different.
*/
int symseq::operator==(symseq y) {
  const long M=y.getN();
  long n;
  int retval;
  char xsym,ysym;
  if(N!=M) {
    fprintf(stderr, "ERROR using symseq x == symseq y operation: size of x (%ld) does not equal size of y (%ld)\n",N,M);
    exit(EXIT_FAILURE);
  }
  for(n=0,retval=1;n<N;n++) {
    xsym= get(n);
    ysym=y.get(n);
    if(xsym!=ysym) retval=0;
  }
  return retval;
}
*/
* external functions
*/
* compare dieseq x and dieseq y
* return the number of locations in which the two sequences are different
* return 0 if the same
*/
long cmp(const symseq *x, const symseq *y, int showdiff, FILE *fptr) {
  const long N=x->getN();
  const long M=y->getN();
  char xsym,ysym;
  long n;
  long count;
  if(N!=M) {
    fprintf(stderr, "ERROR using cmp(symseq x,symseq y): size of x (%ld) != size of y (%ld)\n",N,M);
    exit(EXIT_FAILURE);
  }
  for(n=0,count=0;n<N;n++) {
    xsym=x->get(n);
    ysym=y->get(n);
    if(xsym!=ysym) {
      count++;
      if(showdiff) fprintf(stderr, "%ld: x[%ld]=%c(0x%02x)\n",n,xsym,ym,n,ym,sym,ysym);
      if(fptr!=NULL) fprintf(fptr, "%ld: x[%ld]=%c(0x%02x)\n",n,xsym,ym,n,ym,sym,ysym);
    }
  }
}
F.2 Die routines

```c
/* ==============================================================
 * Daniel J. Greenhoe  
 * header file for routines for die routines 
 * 'A'--> die face value 1  
 * 'B'--> die face value 2  
 * 'C'--> die face value 3  
 * 'D'--> die face value 4  
 * 'E'--> die face value 5  
 * 'F'--> die face value 6 
 */

class dieseq: public symseq { 
    public: 
        dieseq(const long M) : symseq(M) {; // constructor initializing to ' ' 
```
```cpp
dieq(const long M, const unsigned seed) : symseq(M, seed, "ABCDEF") {}; // constructor initializing random values
void randomize(void) {symseq::randomize("ABCDEF");} //
void randomize(unsigned seed) {rand(seed); randomize();}
int randomize(long start, long end, int wa, int wb, int wC, int wD, int we, int wf); int randomize(unsigned seed, int wa, int wb, int wC, int wD, int we, int wf) {rand(seed); return randomize(0, getN() - 1, wa, wb, wC, wD, we, wf);}
int randomize(int wa, int wb, int wC, int wD, int we, int wf) {return randomize(0, getN() - 1, wa, wb, wC, wD, we, wf);} char get(long n) {return symseq::get(n, "ABCDEF");} // get a value from x at location n
void put(long n, char symbol) {symseq::put(n, symbol);} seqR1 dietoR1(void); // map die face values to R^1
deqC1 dietoC1(void); // map die face values to R^1
deqR1 dietoR1pam(void); // map die face values to R^1 using PAM scheme (symmetric about zero)
deqR3 dietoR3(void); // map die face values to R^3
deqR1 histogram(const long start, const long end, int display, FILE *fptr); // compute, display, and write histogram
deqR1 histogram() {return histogram(0, getN() - 1, 0, NULL);} // compute histogram
deqR1 histogram(const long start, const long end) {return histogram(start, end, 0, NULL);} // compute histogram
deqR1 histogram(int display, FILE *fptr) {return histogram(0, getN() - 1, 1, fptr);} // print histogram to file
deqR1 histogram(FILE *fptr) {return histogram(0, getN() - 1, 0, fptr);} // print histogram to file
void operator=(dieseq y); // x=y

extern int die_domain(char c); // check if value is in the domain of die
extern double die_dieoR1 (char c);
extern double die_dieoR1pam (char c);
extern vector3 die_dieoR3 (char c);
extern vector6 die_dieoR6 (char c);
extern complex die_dieoCl (char c);

/*===================================================================*/
* Daniel J. Greenhoe
* routines for Real Die dieq
*===================================================================*/
* headers
*===================================================================*/
#include <stdio.h>
#include <stdlib.h>
#include <string.h>
#include <math.h>
#include <main.h>
#include <sysmsg.h>
#include <r1.h>
#include <r2.h>
#include <r3.h>
#include <r6.h>
#include <c1.h>
#include <die.h>

/*====================================================================*/
* prototypes
*====================================================================*/
void phistogram(seqR1 *data, const long start, const long end, FILE *ptr);

/*====================================================================*/
* fill the dieq with weighted pseudo-random die face values
*====================================================================*/
int dieeq::randomize(long start, long end, int wa, int wb, int wC, int wD, int we, int wf) {
  int r, u;
  long n;
  char symbol;
  int sum=wa*wb*wC*wD*we*wf;
  if(sum=100) {
    printf(stderr, "dieseq::randomize error: sum of weight values = %d != 100\n",sum);
    return -1;
  }
  // print("\nstart=%ld end=%ld weights=\n%03d %03d %03d %03d %03d %03d\n",
      start, end, wa, wb, wC, wD, we, wf);
  for(n=start; n<end; n++){
    r=rand();
    u = r%100;
```
if (u<wA) symbol='A';
else if (u<wA+wB) symbol='B';
else if (u<wA+wB+wC) symbol='C';
else if (u<wA+wB+wC+wD) symbol='D';
else if (u<wA+wB+wC+wD+wE) symbol='E';
else symbol='F';
put(n, symbol);
}
return 0;

/*
* map die face values to R^1
* A-->1 B-->2 C-->3 D-->4 E-->5 F-->6
* all other values --> 0
*/
seqR1 dieseq::diertoR1(void){
    const long N=getN();
    long n;
    char sym;
    double xR1;
    seqR1 y(N);
    for(n=0; n<N; n++){
        sym = get(n);
        xR1 = die_diertoR1(sym);
        y.put(n, xR1);
    }
    return y;
}

/*
* map die face values to R^1 using PM scheme
* A-->2.5 B-->1.5 C-->0.5 D-->0.5 E-->1.5 F-->2.5
* all other values --> 0
*/
seqR1 dieseq::diertoR1pam(void){
    const long N=getN();
    long n;
    char sym;
    double xR1;
    seqR1 y(N);
    for(n=0; n<N; n++){
        sym = get(n);
        xR1 = die_diertoR1pam(sym);
        y.put(n, xR1);
    }
    return y;
}

/*
* map die face values to C^1
*/
seqC1 dieseq::dietoC1(void){
    const long N=getN();
    long n;
    char sym;
    complex xC1;
    seqC1 y(N);
    for(n=0; n<N; n++){
        sym = get(n);
        xC1 = die_dietoC1c(sym);
        y.put(n, xC1);
    }
    return y;
}

/*
* map die face values to R^3 sequence
*/
seqR3 dieseq::diertoR3(void){
    const long N=getN();
    long n;
    char sym;
    vectR3 xR3;
    seqR3 y(N);
    for(n=0; n<N; n++){
        sym = get(n);
        xR3 = die_diertoR3(xR3, sym);
        y.put(n, xR3);
    }
    return y;
}
x3 = die_dietoR3(sym);
y.put(n, xR3);
}
return y;
}

/* compute histogram of dna sequence 
return seqR1 y of length 6 where 
* y[1]--->number of dna 'A' symbols,
* y[2]--->number of dna 'B' symbols,
* y[3]--->number of dna 'C' symbols,
* y[4]--->number of dna 'D' symbols,
* y[5]--->number of dna 'E' symbols,
* y[6]--->number of dna 'F' symbols,
* y[0]--->number of all other values 
y[7]--->total number of symbols y[1], y[2], ..., y[6]
*/
seqR1 dieseq::histogram(const long start, const long end, int display, FILE *fptr) {
seqR1 data(8);
long n;
long bin;
double p;
int i;
char symbol;
FILE *ptr;
data.clear();
for (n=start; n<end; n++) {
    symbol = get(n);
    switch (symbol) {
        case 'A': bin = 1; break;
        case 'B': bin = 2; break;
        case 'C': bin = 3; break;
        case 'D': bin = 4; break;
        case 'E': bin = 5; break;
        case 'F': bin = 6; break;
        default: bin = 0; break;
    }
    if (bin != -1) data.increment(bin);
}
if (display) phistogram(&data, start, end, stdout);
if (fptr != NULL) phistogram(&data, start, end, fptr);
return data;
}

/* print die sequence histogram with data pointed to by <data> 
* to stream pointed to by ptr 
*/
void phistogram(seqR1 *data, const long start, const long end, FILE *ptr) {
const long N = end - start + 1;
long bin;
fprintf(ptr, prn, "\n");
fprintf(ptr, prn, "| Histogram for sequence | n=%ld-%ld | (length %ld) | \n\", start, end, N);
for (bin=1; bin<=6; bin++) fprintf(ptr, prn, " %c\", 'A'+(char)bin-1);
fprintf(ptr, prn, extra, "\n");
for (bin=1; bin<=6; bin++) fprintf(ptr, prn, "%0.0f\", data->get(bin));
for (bin=1; bin<=6; bin++) fprintf(ptr, prn, "%0.1f\", data->get(bin) / (double)N*100.0);
for (bin=1; bin<=6; bin++) fprintf(ptr, prn, "%.2f%%\", data->get(bin) / (double)N*100.0);
fprintf(ptr, prn, extra, "\n");
fprintf(ptr, prn, extra, "\n");
}

/*==============================* /
* operators 
*================================*
/* operator dieseq x = dieseq y */
void dieseq::operator=(dieseq y) {
    const long N = getN();
    const long M = y.getN();
    long n;
    char symbol;

if (N! = M) {
    fprintf(stderr, "ERROR using dieseq x = dieseq y operation: size of x (%ld) does not equal size of
    y (%ld)!n", N, M);
    exit(EXIT_FAILURE);
}
}
for (n = 0; n < N; n++) {
    symbol = y.get(n);
    put(n, symbol);
}
}

/* external operations
  */
/* map die face values to R^1 */
/* DO NOT EDIT THIS CODE */
double die_dieToR1(char c) {
    double rval;
    switch (c) {
        case 'A': rval = 1.0; break;
        case 'B': rval = 2.0; break;
        case 'C': rval = 3.0; break;
        case 'D': rval = 4.0; break;
        case 'E': rval = 5.0; break;
        case 'F': rval = 6.0; break;
        default:
            fprintf(stderr, "ERROR using die_dieToR1(c): c=%c(0x%x) is not in the valid domain
            A,B,C,D,E,F)!n", c, c);
            exit(EXIT_FAILURE);
    }
    return rval;
}

/* map die face values to R^1 PAM */
/* DO NOT EDIT THIS CODE */
double die_dieToR1Pam(char c) {
    double rval;
    switch (c) {
        case 'A': rval = -2.5; break;
        case 'B': rval = -1.5; break;
        case 'C': rval = -0.5; break;
        case 'D': rval = 0.5; break;
        case 'E': rval = 1.5; break;
        case 'F': rval = 2.5; break;
        default:
            fprintf(stderr, "ERROR using die_dieToR1Pam(c): c=%c(0x%x) is not in the valid domain
            A,B,C,D,E,F)!n", c, c);
            exit(EXIT_FAILURE);
    }
    return rval;
}

/* map die face values to complex plane C^1 */
/* DO NOT EDIT THIS CODE */

    imaginary axis
    |
    |
    |
    B=(cos90,sin90)
    |
    |
    (cos150,sin150)=C A=(cos30,sin30)
    |
    real axis
    |
    |
    (cos210,sin210)=D F=(cos330,sin330)
    |
    |
    |
    |
    |
    |
    |
    complex die_dieToClc(char c){
    complex rc;
    switch (c) {
        case 'A': rc = expi( 30.0/180.0*PI); break;
F.3  Real die routines

/*===============================================================*/
/** Daniel J. Greenhoe                                       **
/** header file for routines for real die routines           **
/** 'A'--> die face value 1                                   **
/** 'B'--> die face value 2                                   **
/** 'C'--> die face value 3                                   **
/** 'D'--> die face value 4                                   **
/** 'E'--> die face value 5                                   **
/** 'F'--> die face value 6                                   **

APPENDIX F. C++ SOURCE CODE SUPPORT

case 'B': rc = expi( 90.0/180.0*PI); break;
case 'C': rc = expi(150.0/180.0*PI); break;
case 'D': rc = expi(210.0/180.0*PI); break;
case 'E': rc = expi(270.0/180.0*PI); break;
case 'F': rc = expi(330.0/180.0*PI); break;
    case '0': rc.put(0,0); break;
default: rc.put(0,0);
    fprintf(stderr, "ERROR using dietoC1 (char c): c=%c(0x%x) is not in the valid domain
                   \{0,A,B,C,D,E,F\}. Returning (0,0).\n",c,c);
 }
return rc;
}

/*===============================================================*/
* map die face values to R^3
* +1 |  0 |  0 |  0 |  0 |  0 |  0 |  0 | -1 |
* 0 |  0 |  0 |  0 |  0 |  0 |  0 |  0 |  0 |
*---------------------------------------------------------------*/
    vect3 die_diotoR3(char c){
        vect3 xzy;
        switch(c){
            case 'A': xzy.put(+1, 0, 0); break;
            case 'B': xzy.put(+1, +1, 0); break;
            case 'C': xzy.put( 0, +1, +1); break;
            case 'D': xzy.put( 0,  0, +1); break;
            case 'E': xzy.put( 0,  0,  0); break;
            case 'F': xzy.put(+1,  0,  0); break;
        default:
            fprintf(stderr, "ERROR using die_diotoR3 (c): c=%c(0x%x) is not in the valid domain
                   \{A,B,C,D,E,F\}.\n",c,c);
            exit(EXIT_FAILURE);
        }
        return xzy;
    }

/*===============================================================*/
* map die face values to R^6
* +1,0,0,0,0,0,0 | 0,0,0,1,0,0 | 0,1,0,0,0,0 | 0,0,1,0,0,0 | 0,0,0,0,1,0 | 0,0,0,0,0,1 |
* 0 -->(0,0,0,0,0,0) |
*---------------------------------------------------------------*/
    vect6 die_diotoR6(char c){
        vect6 rsix;
        switch(c){
            case 'A': rsix.put(1, 0, 0, 0, 0, 0); break;
            case 'B': rsix.put(0,1, 0, 0, 0, 0); break;
            case 'C': rsix.put(0,0, 1, 0, 0, 0); break;
            case 'D': rsix.put(0,0, 0, 0, 0, 1); break;
            case 'E': rsix.put(0,0, 0, 0, 0, 0); break;
            case 'F': rsix.put(0,0, 0, 0, 0, 0); break;
        default:
            fprintf(stderr, "ERROR using die_diotoR6 (char c): c=%c(0x%x) is not in the valid domain
                   \{0,A,B,C,D,E,F\}.\n",c,c);
            exit(EXIT_FAILURE);
        }
        return rsix;
    }

}
class rdieseq: public dseq {
  public:
  rdieseq(const long M) : dseq(M) {};
  rdieseq(const long M, const unsigned seed) : dseq(M, seed) {};
  void operator+(dieseq y); // x=y
  void operator>>(=long n) | shiftR(n); // shift rdieseq n elements to the right
  void operator<<(=long n) | shiftL(n); // shift rdieseq n elements to the left
  int metrictbl(void);
  int Rxx(const seqR1 *rxx, const int showcound);
  int Rxx(const seqR1 *rxy, const int showcound, const long N, const long M, const long start, const long finish);
  int Rxxo(const seqR1 *rxx, const int showcound);
  double Rxx(const long m);
};
extern rdieseq rdie_R1todie_euclid(seqR1 xyz);
extern rdieseq rdie_R3todie_larc(seqR3 xyz);
extern rdieseq rdie_R3todie_larc0(seqR3 xyz);
extern rdieseq rdie_R3todie_euclid(seqR3 xyz);
extern rdieseq rdie_R3todie_euclid(seqR3 xyz);
extern vecR3 rdie_dietoR3(char c);
extern int rdie_dietor1(char c);
extern int rdie_domain(char c); // check if value is in the domain of rdie
extern double rdie_metric(char a, char b);
extern double rdie_metric(rdieseq x, rdieseq y); // metric for two sequences

/*==============================================*/
/* Daniel J. Greenhoe
* routines for Real Die rdieseqs
*==============================================*/
/*==============================================*/
/* headers
*==============================================*/
#include<stdio.h>
#include<stdlib.h>
#include<math.h>
#include<iostream.h>
#include<r1.h>
#include<r2.h>
#include<r3.h>
#include<r4.h>
#include<r6.h>
#include<r1h.h>
#include<euclid.h>
#include<larc.h>
#include<rdie.h>
#include<realdie.h>

/* display real die metric table
*==============================================*/
int rdieseq::metrictbl( void ) {
  char a,b;
  for( a='A'; a<='F'; a++ ) {
    for( b='A'; b<='F'; b++ ) {
      printf( "d(\%c,\%c)=\%.1lf \n", a,b, rdie_metric(a,b));
      printf( "\n" );
    }
    return 1;
  }
}

/* autocorrelation Rxx of a real die seqR1 x with 2N offset
*==============================================*/
int rdieseq::Rxxo(const seqR1 *rxx, const int showcound) {
  const long N=getN();
  int rval;
  rval=Rxx(rxx, showcound);
  rxx->add(2*N);
  return rval;
}

/* autocorrelation Rxx of a real die seqR1 x
*==============================================*/
int rdieseq::Rxx(const seqR1 *rxx, const int showcound) {

long m;
cost long N=getN();
\textbf{int rval=0;}
double rxm;
if(showcount) printf(stderr," Calculate %ld auto-correlation values ... n="\n2*N+1);
for(m=N;m>0;m++){
    if(showcount) printf(stderr,"%ld m=N);
    rxm=Rxx(m);
    if(rxm>0) rval=-1;
    rxm->put(m,N,rxm);
    if(showcount) printf(stderr,"\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b")
}
if(showcount) printf(stderr,"%ld .... done.\n",mN);
return rval;
}

/*-------------------------------------------------------------
 * autocorrelation Rxx(m)
 *-------------------------------------------------------------*/
double rdieseq::Rxx(const long m){
    const long mm=labs(m);
    cost long N=getN();
    long n,mm;
    double d,sum;
    char a,b;
    for(n=0;sum=0;N<0(N=mm);n++){
        mm=n-mm;
        a=(n < 0 || n >=N)? 0.0 : get(n);
        b=(mm=0 || mm=N)? 0.0 : get(mm);
        d=(a==0 || b==0)? 1.0 : rdie_metric(a,b);
        sum+=d;
    }
    return -sum;
}

/*==================================================================
 * operators
 *==================================================================*/
/*==================================================================
 * operator rdieseq x = diseq y
 *==================================================================*/
void rdieseq::operator=(dieseq y){
    long n;
    cost long N=getN();
    cost long M=y.getN();
    char symbol;
    \textbf{if}(N>M) {
        printf(stderr,"ERROR using rdieseq x = diseq y operation: size of x (%ld) does not equal size of y (%ld)\n",N,M);
        exit(EXIT_FAILURE);
    }
    for(n=0;n<N;n++){
        symbol=y.get(n);
        put(n,symbol);
    }
}

/*==================================================================
 * external operations
 *==================================================================*/
/*==================================================================
 * map die face values to R^1
 *==================================================================*/
* A-->1 B-->2 C-->3 D-->4 E-->5 F-->6

\textbf{int} rdie_dietoR1(char c){
    \textbf{int} n, rval;
    char domain[6]=\{\textquoteleftA\textquoteright, \textquoteleftB\textquoteright, \textquoteleftC\textquoteright, \textquoteleftD\textquoteright, \textquoteleftE\textquoteright, \textquoteleftF\textquoteright\};
    char element;
    for(n=0,rval=-1;n<6;n++)if(c==domain[n]) rval=n;
    if(rval=-1){
        printf(stderr,"ERROR using rdie_dietoR1(char c): c=%c(0x%x) is not in the valid domain 0,A,B,C,D,E,F)\n",c,c);
        exit(EXIT_FAILURE);
    }
    return rval;
}
/* map die face values to RA^3 */

* |+1| | 0| | 0| | 0| | 0| | 0| | 0| | 0| | 0| | 0|
* A->>0 B->>0 C->>0 D->>0 E->>-1 F->>0 0->>0

/*

```c
vectR3 rdie_dietoR3(char c)
{
    vectR3 xyz;
    switch(c)
    {
        case 'A': xyz.put(+1, 0, 0); break;
        case 'B': xyz.put( 0, 0, 1); break;
        case 'C': xyz.put( 0, 0,-1); break;
        case 'D': xyz.put(+1, 0, 0); break;
        case 'E': xyz.put( 0,-1, 0); break;
        case 'F': xyz.put(-1, 0, 0); break;
        default:
            fprintf(stderr,"ERROR using rdie_dietoR3(char c): c=%c(0x%x) is not in the valid domain
 [A,B,C,D,E,F]\n",c,c);
            exit(EXIT_FAILURE);
    }
    return xyz;
}
```
*/

/* map RA^3 values to die face values using Lagrange Arc metric */

```c
rdieseq rdie_R3todie_larc(seqR3 xyz)
{
    long n;
    int m;
    long N=xyz.getN();
    double d[7];
    double smallestd;
    char closestface;
    vectR3 p,q[7];
    rdieseq rdie(N);
    //q[0].put(0,0,0);
    q[1]=rdie_dietoR3('A');
    q[2]=rdie_dietoR3('B');
    q[3]=rdie_dietoR3('C');
    q[4]=rdie_dietoR3('D');
    q[5]=rdie_dietoR3('E');
    q[6]=rdie_dietoR3('F');
    for(n=0; n<N; n++)
    {
        p.put(xyz.getx(n),xyz.gety(n),xyz.getz(n));
        smallestd=larc_metric(p,q[1]);
        closestface='A';
        for(m=2;m<7;m++)
        {
            d[m]=larc_metric(p,q[m]);
            if( ((m&0x01) && (d[m]<smallestd)) || (!((m&0x01)) && (d[m]<=smallestd)) )
            /
            // bias odd samples
            // towards smaller values
            smallestd=d[m];
            closestface='A'*m-1;
        }
        rdie.put(n,closestface);
    }
    return rdie;
}
```

/* map RA^3 values to die face values and (0,0,0) using Lagrange Arc metric */

```c
rdieseq rdie_R3todie0_larc(seqR3 xyz)
{
    long n;
    int m;
    long N=xyz.getN();
    double d[7];
    double smallestd;
    char closestface;
    vectR3 p,q[7];
    rdieseq rdie(N);
    q[0].put(0,0,0);
```
q[1] = rdie_dieterR3('A');
q[2] = rdie_dieterR3('B');
q[3] = rdie_dieterR3('C');
q[4] = rdie_dieterR3('D');
q[5] = rdie_dieterR3('E');
q[6] = rdie_dieterR3('F');

for (n=0; n<N; n++) {
p.put(xyz.getn(), xyz.getn(), xyz.getn());
smallest = larc_metric(p, q[0]);
// smallest = ae_metric(1, p, q[0]);
closestface = '0';
for (m=1; m<=7; m++) {
d[m] = larc_metric(p, q[m]);
if ( ((n&0x01) & (d[m]<smallest)) || (!!(n&0x01) & (d[m]<smallest))) {
  // bias odd samples bias even samples
  // bias towards smaller values towards larger values
  smallest = d[m];
  closestface = 'A'+m-1;
}
rdie.put(n, closestface);
}
return rdie;
}

/*-----------------------------*/
* map R^3 values to die face values using Euclidean metric
* 0  A  B  C  D  E  F  A...+F
*-----------------------------*/
rdieseq rdie_R3todie_euclid (seqR3 xyz) {
  long n;
  int m;
  long N=xyz.getN();
  double d[7];
  double smallest;
  char closestface;
  vectR3 p, q[7];
  rdieseq rdie (N);
  q[1] = rdie_dieterR3('A');
  q[2] = rdie_dieterR3('B');
  q[3] = rdie_dieterR3('C');
  q[4] = rdie_dieterR3('D');
  q[5] = rdie_dieterR3('E');
  q[6] = rdie_dieterR3('F');
  for (n=0; n<N; n++) {
p.put(xyz.getn(), xyz.getn(), xyz.getn());
smallest = ae_metric(1, p, q[1]);
closestface = 'A';
for (m=2; m<=6; m++) {
d[m] = ae_metric(1, p, q[m]);
if ( ((n&0x01) & (d[m]<smallest)) || (!!(n&0x01) & (d[m]<smallest))) {
  // bias towards smaller values bias towards larger values
  smallest = d[m];
  closestface = 'A'+m-1;
}
rdie.put(n, closestface);
}
return rdie;
}

/*-----------------------------*/
* map R^3 values to die face and (0,0,0) values using Euclidean metric
* 0  A  B  C  D  E  F  A...+F
*-----------------------------*/
rdieseq rdie_R3todie0_euclid (seqR3 xyz) {
  long n;
  int m;
  long N=xyz.getN();
  double d[7];
  double smallest;
  char closestface;
  for (n=0; n<N; n++) {
p.put(xyz.getn(), xyz.getn(), xyz.getn());
smallest = larc_metric(p, q[0]);
// smallest = ae_metric(1, p, q[0]);
closestface = '0';
for (m=1; m<7; m++) {
d[m] = larc_metric(p, q[m]);
if ( ((n&0x01) & (d[m]<smallest)) || (!!(n&0x01) & (d[m]<smallest))) {
  // bias odd samples bias even samples
  // bias towards smaller values towards larger values
  smallest = d[m];
  closestface = 'A'+m-1;
}
rdie.put(n, closestface);
}
return rdie;
}
vector3 p, q[7];
rdieseq rdie(N);

def Flores rdie_dietoR3('0');
q[0]=rdie_dietoR3('A');
q[1]=rdie_dietoR3('B');
q[2]=rdie_dietoR3('A');
q[3]=rdie_dietoR3('C');
q[4]=rdie_dietoR3('D');
q[5]=rdie_dietoR3('E');
q[6]=rdie_dietoR3('F');

for(n=0; n<N; n++) {
    p.put((xyz.getx(n),xyz.gety(n),xyz.getz(n));
    smallestd=ae_metric(1,p,q[0]);
    closestface='0';
    for(m=1;m<=6;m++) {
        d[m]=ae_metric(1,p,q[m]);
        if(d[m]<smallestd)
            // bias towards smaller values
            closestface='A' + m-1;
    }
    // if(m==0) /* (alternative coding)
    // smallestd=d[m];
    // closestface='A' + m-1;
    // */
    // else
    //     if(d[m]<smallestd)
    //         // smallestd=d[m];
    //         // closestface='A' + m-1;
    //     //
    // }
rdie.put(n,closestface);
}
return rdie;

/* map R^3 values to die face values using Euclidean metric */
rdieseq rdie_R1todie_euclid(seqR1 xyz){
    long n;
    long N=xyz.getN();
    char closestface;
    double p;
    rdieseq rdie(N);

    for(n=0; n<N; n++) {
        p=xyz.get(n);
        if(p<1.5) closestface='A';
        else if(p>=5.5) closestface='F';
        else closestface=(char)(p*0.5-1)+'A';
rdie.put(n,closestface);
    }
return rdie;
}

/* real die metric d(a,b) */
* d(a,b) | 0  A  B  C  D  E  F
* a= | 0  1  1  1  1  1  1
* a= | A  1  1  1  1  1  2
* a= | B  1  1  0  1  1  2 1
* a= | C  1  1  1  0  2  1  1
* a= | D  1  1  1  2  0  1  1
* a= | E  1  1  2  1  1  0  1
* a= | F  1  2  1  1  1  1  0
* On success return d(a,b). On error return -1.
*/
double rdie_metric(char a, char b){
    /*
*/
}
F.4 Spinner routines

```c
/*..............................................................................................*/
class spinseq: public dieseq {
    public:
        spinseq(const long M) : dieseq(M) {};
        spinseq(const long M,const unsigned seed) : dieseq(M,seed) {};
        seq1 spintoR1(void);    //map spin face values to R^1
        seq2 spintoR2(void);    //map spin face values to R^2
        void operator=(spinseq y); //x=y
        int metricR1(void);
        double Rxx (const long m);
        int Rxx (const seq1 *rxx, const int showcount);
        int Rxxo(const seq1 *rxx, const int showcount);
        spinseq downsamp1(int factor);//downsample by a factor of <factor>
    };
    extern int spin_domain(char c);//check if value is in the domain of rspin
    extern spinseq spin_R1tospin_euclid(seq1 xy);
    extern spinseq spin_R2tospin_larc(seq2 xy);
    extern spinseq spin_R2tospin0_larc(seq2 xy);

    /***********************************************************/
    * real die metric p(x,y) where x and y are rdie sequences computed as
    * p(x,y) = d(x0,y0) + d(x1,y1) + d(x2,y2) + ... + d(x[N-1],y(N-1))
    * where d(a,b) is defined above.
    * On success return d(x,y). On error return -1.
    */
    double rdie_metric(rdie seq x, rdie seq y){
        double rval,d;
        long n;
        long N=x.getN();
        long M=y.getN();
        long NM=(N*M)/M:N;  //NM = the larger of N and M
        for(n=0,d=0;n<NM;n++){
            rval=rdie_metric(x.get(n),y.get(n));
            if(rval<0){d+=0.0; printf("%f ",rval);
            else d+=rval;
        }
        if(N!=M){
            printf(stderr,"ERROR using rdie_metric(rdie seq x,rdie seq y): size of x (%ld) does not equal the size of y (%ld).\n",N,M);
            exit(EXIT_FAILURE);
        }
        return d;
    }
};
*/
```

**APPENDIX F. C++ SOURCE CODE SUPPORT**

int ra=rdie_dietoR1(a);
int rb=rdie_dietoR1(b);
double d;

if(ra<0) {printf(stderr,"%s\n",a); d=1;
else if(rb<0) d=-1;
else if(ra==rb) {d=0; }
else if(ra==0) d=1;
else if(rb==0) d=1;
else if(ra+rb==7) d=2;
else d=1;
return d;
}

*real* die metric p(x,y) where x and y are rdie sequences computed as
p(x,y) = d(x0,y0) + d(x1,y1) + d(x2,y2) + ... + d(x[N-1],y(N-1))
where d(a,b) is defined above.
On success return d(x,y). On error return -1.

---

Daniel J. Greenhoe

---

A thesis concerning symbolic sequence processing

Daniel J. Greenhoe

Version 0.50E
extern spinseq spin_R2tospin_euclid(seqR2 xy);
extern spinseq spin_R2tospin0_euclid(seqR2 xy);
extern vecR2 spin_spintoR2(char c);
extern double spin_spintoR1(char c);
extern double spin_metric(char a, char b);
extern double spin_metric(spinseq x, spinseq y); // metric for two sequences
//extern seqR1 spin_correlation(seqseq x, seqseq y, int showcount); // correlation
//extern seqR1 spin_correlation(seqseq x, seqseq y) return spin_correlation(x,y,0); // correlation

/*==================================*/
/* routines for Real spin seqseqs
/*==================================*/
/*==================================*/
/* display spin metric table        */
/*==================================*/
int spinseq :: metrictbl(void)
char a,b;
for(a=A; a=F; a++)
  for(b=A; b=F; b++) printf("d(%c,%c)=%.1f ",a,b,spin_metric(a,b));
  printf("\n");
return 1;
/*==================================*/
/* autocorrelation Rxx of a spinseq with 2N offset */
/*==================================*/
int spinseq :: Rxxo(const seqR1 *rxx, const int showcount)
const long N=getN();
int rval;
  rval=Rxx(rxx,showcount);
  rxx->add(2*N);
  return rval;
/*==================================*/
/* autocorrelation Rxx of a spinseq */
/*==================================*/
int spinseq :: Rxx(const seqR1 *rxx, const int showcount)
long m;
const long N=getN();
int rval=0;
double rxxm;
if(showcount) printf(stderr," Calculate %ld auto-correlation values ... n=%",2*N+1);
  for(m=N;m<N;m++)
    if(showcount) printf(stderr,"%8ld","m=N);
    rxxm=Rxx(m);
    if(rxxm>0) rval--;  
    rxx=put(m,N,rxxm);
    if(showcount) printf(stderr,"\b\b\b\b\b\b\b\b\b\b\b\b\b\b");
  }
  if(showcount) printf(stderr,"\n","m=N);
  return rval;
  }
/*==================================*/
/* autocorrelation Rxx(m)        */
double spinseq::Rxx(const long m)
{
    const long mm= labs(m);
    const long N=getN();
    long n,mm;
    double d,sum;
    char a,b;
    for(n=0; sum=0; n<(N=mm); n++)
    {
        mm=n-mm;
        a=(n < 0 || n >=N)? 0.0 : get(n);
        b=(mm<0 || mm=N)? 0.0 : get(mm);
        d=(a==0 || b==0)? 1.0 : spin_metric(a,b);
        sum+=d;
    }
    return -sum;
}

* downsample sequence by a factor of <factor>*

spinseq spinseq::downsample(int factor)
{
    const long N=getN();
    long n,m;
    long M;
    if(factor<1)
    {
        printf(stderr,"ERROR using die_seq::downsample: factor=%d must be at least 1\n",factor);
        exit(EXIT_FAILURE);
    }
    M=N/factor;
    spinseq newseq(M);
    for(n=0; m=0; m=M; n+=factor ,m++)newseq.put(m, get(n));
    return newseq;
}

* map spin face values to R^1
* A->1 B->2 C->3 D->4 E->5 F->6

seqR1 spinseq::spintoR1(void)
{
    const long N=getN();
    long n;
    seqR1 y(N);
    for(n=0; n<N; n++)y.put(n, spin_spintoR1(get(n)));
    return y;
}

* map spin face values to R^2 sequence

seqR2 spinseq::spintoR2(void)
{
    const long N=getN();
    long n;
    seqR2 seqR2(N);
    for(n=0; n<N; n++)seqR2.put(n, spin_spintoR2(get(n)));
    return seqR2;
}

* operators

void spinseq::operator = (spinseq y)
{
    long n;
    const long N=getN();
    long M=y.getN();
    char symbol;
    if(N>M){
        printf(stderr,"\nERROR using spinseq x = spinseq y: size of x (%Id) is smaller than size of y (%Id) \n",N,M,N);
        exit(EXIT_FAILURE);
    }
    for(n=0; n<N; n++)
   symbol=y.get(n);
    // towards smaller values towards larger values
    smallestd=d[m];
    closestface='A'+m-1;
  }
  rspin.put(n,closestface);
}
return rspin;
}

/*
 * map R^3 values to spin face values and (0,0,0) using Lagrange Arc metric
 */
spinseq spin_R2tospin0_larc(seqR2 xy){
  long n;
  int m;
  long N=xy.getN();
  double d[7];
  double smallestd;
  char closestface;
  vectR2 p,q[7];
  spinseq rspin(N);
  q[0],put(0,0);
  q[1]=spin_spintoR2('A');
  q[2]=spin_spintoR2('B');
  q[3]=spin_spintoR2('C');
  q[4]=spin_spintoR2('D');
  q[5]=spin_spintoR2('E');
  q[6]=spin_spintoR2('F');
  for(n=0; n<N; n++){
    p.put(xy.getx(n),xy.gety(n));
    smallestd=larc_metric(p,q[0]);
    //smallestd=ae_metric(1,p,q[0]);
    closestface='0';
    for(m=1;m<=7;m++)
      d[m] = larc_metric(p,q[m]);
    if(((m&0x01) && (d[m]<smallestd)) || (!((m&0x01)) && (d[m]<smallestd)))
      // bias odd samples bias even samples
      // towards smaller values towards larger values
      closestface='A'+m-1;
  }
  rspin.put(n,closestface);
}
return rspin;
}

/*
 * map R^2 values to spin face values using Euclidean metric
 */
spinseq spin_R2tospin_euclid(seqR2 xy){
  long n;
  int m;
  long N=xy.getN();
  double d[7];
  double smallestd;
  char closestface;
  vectR2 p,q[7];
  spinseq rspin(N);
  q[1]=spin_spintoR2('A');
  q[2]=spin_spintoR2('B');
  q[3]=spin_spintoR2('C');
  q[4]=spin_spintoR2('D');
  q[5]=spin_spintoR2('E');
  q[6]=spin_spintoR2('F');
  for(n=0; n<N; n++){
    p.put(xy.getx(n),xy.gety(n));
    smallestd=ae_metric(1,p,q[1]);
    closestface='A';
    for(m=2;m<=7;m++)
      ...
d[m] = ae_metric(1,p,q[m]);
if(((m&0x01) && (d[m]<smallest)) ||
    ((!(m&0x01)) && (d[m]<=smallest))){
    // bias towards smaller values
    smallest=d[m];
    closestface='A'+m-1;
}
}
s pin . put(n, closestface);
}
return rpin;

/*
 * map R^2 values to spin face and (0,0) values using Euclidean metric
 */
spinseq spin_R2spin0_euclid(seqR3 xy){
    long n;
    int m;
    long N=xy.getN();
double d[7];
double smallest;
char closestface;
vecR2 p,q[7];
spinseq rpin(N);
    q[0]=spin_spintoR2('0');
    q[1]=spin_spintoR2('A');
    q[2]=spin_spintoR2('B');
    q[3]=spin_spintoR2('C');
    q[4]=spin_spintoR2('D');
    q[5]=spin_spintoR2('E');
    q[6]=spin_spintoR2('F');
    for(n=0; n<N; n++) {
        p.put(xy.getx(n),xy.gety(n));
        smallest=ae_metric(1,p,q[0]);
        closestface='0';
        for(m=1;m<=6;m++)
            if(d[m]<smallest)
                if((!(m&0x01)) && (d[m]<=smallest))
                    // bias towards smaller values
                    smallest=d[m];
                    closestface='A'+m-1;
            } else
                if((!(m&0x01)) && (d[m]<smallest))
                    // bias towards larger values
                    smallest=d[m];
                    closestface='A'+m-1;
            }
    }
    return rpin;
}

/*
 * map R^1 values to spin face values using Euclidean metric
 */
spinseq spin_R1spin_euclid(seqR1 xy){
    long n;
    long N=xy.getN();
char closestface;
double p;
spinseq rpin(N);
}
E.5 DNA routines

```cpp
/*=============================================*/
* Daniel J. Greenhoe
* header file for routines for DNA routines
*=============================================*/

class dnaseq { public:

dnaseq(long M) : symseq[M]{}; // constructor initializing to '.
void seed(unsigned seed) {rand.seed();} //
void randomize(void) {symseq::randomize("ATCG");} //
void randomize(unsigned seed) {rand.seed(); randomize();}
int randomize(long start, long end, int wa, int wT, int wC, int wG); //
int randomize(unsigned seed, int wa, int wT, int wC, int wG) { return randomize(0, getN() - 1, wa, wT, wC, wG); }
int randomize(long start, long end, unsigned seed, int wa, int wT, int wC, int wG) { return randomize(0, getN() - 1, wa, wT, wC, wG); }
char get(long n) { return symseq::get(n,"ATCG"); } // get a value from x at location n
void put(long n, char symbol) {symseq::put(n,symbol);} //
void put(dnaseq *y, const long n, const char symbol);
void put(dnaseq *y, long n) {return put(y,n,'.');} //
void put(const long start, const long end, char c); // put a value <c> at locations start to end
seqR2 dnaR2(void){} // map gsp face values to R^2
seqR2 downsample(int factor) { // downsample by a factor of <factor>
seqR1 dnaR1(void){} // map dna sequence to R^1
seqC1 dnaC1(void){} // map dna sequence to R^1
seqR1 dnaR1pam(void); // map dna sequence to R^1 using Pam scheme (symmetric about zero)
seqR1 dnaR1bin(void); // map dna sequence to R^1 using AT-CG binary scheme
seqR4 dnaR4(void){} //map dna sequence to R^4
double Rxx (const long m);
int Rxx (const seqR1 *Rxx, const int showcount);
int Rxx0(const seqR1 *rxx, const int showcount);
seqR1 histogram(const long start, const long end, int display, FILE *fptr); // compute, display, and write histogram
seqR1 histogram() { return histogram(0, getN() - 1,0, NULL); } // compute histogram
seqR1 histogram(const long start, const long end) { return histogram(start, end, 0, NULL); } // compute histogram
```
```c
/*========================================================*/
/*    * Daniel J. Greenhoe
/*    * routines for Real gsp dnaseqs
/*    *========================================================*/
#include<stdio.h>
#include<stdlib.h>
#include<string.h>
#include<math.h>
#include<sysseq.h>
#include<r1.h>
#include<r2.h>
#include<r3.h>
#include<r4.h>
#include<r6.h>
#include<r1.c>
#include<euclid.h>
#include<larc.h>
#include<dna.h>

/* prototypes
*========================================================*/
void dna_phistogram(seqRl *data, const long start, const long end, FILE *ptr);

/* copy a dnaseq <y> into dnaseq x starting at location <n>
* and fill any remaining locations with <c>
*========================================================*/
void dnaseq::put(dnaseq *y, const long n, const char symbol){
    const long N=getN();
    long i,j;
    long M=y->getN();
    if (n>N){
        fprintf(stderr, "nERROR using dnaseq::put(y,n,symbol): n=%ld outside sequence domain
                   [0:%ld]\n",n,N-1);
        exit(EXIT_FAILURE);
    }
    if (!dna_domain(symbol)) {
        fprintf(stderr, "nERROR using dnaseq::put(y,n,symbol): symbol=""%c""=0x%hx not in sequence range
                   [A,T,C,G]\n",symbol);
        exit(EXIT_FAILURE);
    }
    else{
        for (i=0;i<n;i++) put(i,symbol);
        for (j=0;j<M;j++,
             i+=1) put(i,y->get(j));
        for ( ;i<N;i++) put(i,symbol);
    }
}

/* put the value <c> into the sequence x from location <start> to <end>
*========================================================*/
void dnaseq::put(const long start, const long end, const char symbol){
    const long N=getN();
    long n;
    if (start<0||end>N) start=end;
    fprintf(stderr, "nERROR using dnaseq::put(%ld,%ld,'%c')\n",start,end,symbol);
    exit(EXIT_FAILURE);
}
```

**Page 200 of 300**
```cpp
for (n=start; n<=end; n++) put(n, symbol);
}

/* fill the dnaseq with pseudo-random DNA values
 * using seed value <seed>
 * distributed with the weight values <wA,wT,wC,wG>
 * where each weight value wX in an integer in the closed interval [0,100]
 * and where the sum of the intervals must be 100.
 */
dnaseq::randomize(long start, long finish, int wA, int wT, int wC, int wG)
{
    int r,u;
    long n;
    int sum=wA+wT+wC+wG;
    char symbol;
    if (sum!=100)
        fprintf(stderr,"ERROR using dnaseq::randomize(start,finish,wA,wT,wC,wG): sum of weight values != 100\n",sum);
    exit(EXIT_FAILURE);
    for (n=start; n<=finish; n++)
    { r = rand();
        u = r%100;
        if (u<wA) symbol='A';
        else if (u<wA+wT) symbol='T';
        else if (u<wA+wT+wC) symbol='C';
        else symbol='G';
        put(n,symbol);
    }
    return 0;
}

/* map dna face values to R^3 using PAM scheme
 * A-->1.5 T-->3.0 C-->0.5 G-->0.5
 * all other values --> 0
 */
dnaseq::dnatra0lpam(void)
{
    const long N=getN();
    long n;
    char symbol;
    dnaR1 seqR1(N);
    for (n=0; n<N; n++)
    { symbol=get(n);
        switch(symbol)
        { case 'A': seqR1.put(n,-1.5); break;
            case 'C': seqR1.put(n,-0.5); break;
            case 'G': seqR1.put(n, 1.5); break;
            default: fprintf(stderr,"\nERROR using dnaseq::dnatra0lpam(): symbol='x' not in sequence range [A,T,C,G]\n",symbol,symbol);
                exit(EXIT_FAILURE);
        }
    }
    return seqR1;
}

/* map dna face values to R^3 using AT/CG binary scheme
 * A-->1 T-->1 C-->1 G-->1
 * all other values --> 0
 */
dnaseq::dnatra0lbinary(void)
{
    const long N=getN();
    long n;
    char symbol;
    dnaR1 seqR1(N);
    for (n=0; n<N; n++)
    { symbol=get(n);
        switch(symbol)
        { case 'A': seqR1.put(n, 1); break;
            case 'C': seqR1.put(n,-1); break;
            case 'T': seqR1.put(n, 1); break;
            case 'G': seqR1.put(n,-1); break;
            default :
```
```c
fprintf(stderr, "\nERROR using dna2seq:dnaR1bin(): symbol='\%c' not in sequence range [A,T,C,G]\n\n', symbol, symbol);
    exit(EXIT_FAILURE);
}
return seqR1;
}

/* map dna face values to C^1
   -----------------------------------------------*/
seqC1 dna2seq::dnaC1(void)
{
    const long N=getN();
    long n;
    char symbol;
    complex yy;
    seqC1 y[N];
    for(n=0; n<N; n++)
    {
        symbol = get(n);
        yy = dnaC1c(symbol);
        y.put(n,yy);
    }
    return y;
}

/* map dna face values to complex plane C^1
   -----------------------------------------------
   imaginary axis
   -----------------------------------------------
   * (cos135, sin135)\n   * A=(cos45, sin45)
   * --------- real axis
   * (cos225, sin225)\n   * G=(cos315, sin315)
   * 
   * 
   * 
   * 
   */
complex dna2C1c(char c){
    complex rc;
    switch(c){
    case 'A': rc = expi( 45.0/180.0*PI) ; break;
    case 'C': rc = expi(135.0/180.0*PI) ; break;
    case 'T': rc = expi(225.0/180.0*PI) ; break;
    case 'G': rc = expi(315.0/180.0*PI) ; break;
    // case 'A': rc = expi( 45.0/180.0*PI) ; break;  // Gilles 2007 mapping
    // case 'G': rc = expi(135.0/180.0*PI) ; break;
    // case 'T': rc = expi(225.0/180.0*PI) ; break;
    // case 'A': rc = expi(315.0/180.0*PI) ; break;
    // case 'A': rc.put(+1, +1); break;
    // case 'G': rc.put(-1, +1); break;
    // case 'T': rc.put(+1, -1); break;
    // case 'A': rc.put(+1, -1); break;
    // default: rc.put( 0, 0); break;
    printf(stderr, "ERROR using dna2C1c(char c): c='%c' not in the valid domain \n\n', symbol, symbol);
    }
    return rc;
}

/* map dna values to R^4 sequence
   -----------------------------------------------*/
seqR4 dna2seq::dnaR4(void)
{
    const long N=getN();
    long n;
    char yc;
    seqR4 seq4(N);
    for(n=0; n<N; n++)seq4.put(n, dna2R4c(get(n)));
    return seq4;

```
* map dna face values to R^4
  * A->(1,0,0,0)  T->(0,0,1,0)
  * C->(0,1,0,0)  G->(0,0,0,1)
  * 0->(0,0,0,0)
  * on ERROR return (0,0,0,0)

```cpp
*/

void dnaToR(vector<double> &v, char c) {  
  switch(c) {  
    case 'A':  v[0] = 1;  break;  
    case 'T':  v[1] = 1;  break;  
    case 'G':  v[2] = 1;  break;  
    case 'C':  v[3] = 1;  break;  
  }  
}

*/

* map gsp face values to R^3
  * A->1  T->2  C->3  G->4

```cpp
*/

void seq1::dnaToR1(vector<int> &v, int c) {  
  switch(c) {  
    case 'A':  v[0] = 1;  break;  
    case 'T':  v[1] = 1;  break;  
    case 'G':  v[2] = 1;  break;  
    case 'C':  v[3] = 1;  break;  
  }  
}

*/

* map gsp face values to R^2 sequence

```cpp
*/

void seq2::dnaToR2(vector<int> &v, int c) {  
  switch(c) {  
    case 'A':  v[0] = 1;  break;  
    case 'T':  v[1] = 1;  break;  
    case 'G':  v[2] = 1;  break;  
    case 'C':  v[3] = 1;  break;  
  }  
}

*/

* downsample seq1 by a factor of <factor>

```cpp
*/

dnaToR seq1::downsample(int factor) {  
  const long N=getN();  
  long n;  
  char symbol;  
  seq1 seq1(N);  
  for(n=0; n<N; n++) {  
    symbol = get(n);  
    switch(symbol) {  
      case 'A':  seq1.put(n,1);  break;  
      case 'T':  seq1.put(n,2);  break;  
      case 'G':  seq1.put(n,3);  break;  
      case 'C':  seq1.put(n,4);  break;  
      default:  seq1.put(n,0);  break;  
    }  
  }  
  return seq1;  
}

*/

* downsample seq2 by a factor of <factor>

```cpp
*/

dnaToR seq2::downsample(int factor) {  
  const long N=getN();  
  long n;  
  char symbol;  
  seq2 seq2(N);  
  for(n=0; n<N; n++) {  
    symbol = get(n);  
    seq2.put(n,symbol);  
  }  
  return seq2;  
}

*/

fprintfp(stderr, "\nERROR using dnaseq::downsample: factor=%d must be at least 1\n", factor);
exit(EXIT_FAILURE);

M=N/factor;
dnaseq newseq(M);
for(n=0,m=0; m<M; n+=factor,m++)
symbol=get(n);
newseq.put(m,symbol);
return newseq;
}

/*
* compute histogram of dna sequence
* return seq1 y of length 6 where
* y[1]--number of dna 'A' symbols,
* y[2]--number of dna 'T' symbols,
* y[3]--number of dna 'C' symbols,
* y[4]--number of dna 'G' symbols,
* y[5]--number of all other values
* y[5]--total number of symbols y[1], y[2], ..., y[5]
*/

seq1 dnaseq::histogram(const long start, const long end, int display, FILE *fptr){
seq1 data(6);
long n;
long bin;
double p;
int i;
char symbol;
FILE *ptr;
data.clear();
for(n=start;n<=end;n++)
symbol=get(n);
switch(symbol){
case 'A': bin=1; break;
case 'T': bin=2; break;
case 'C': bin=3; break;
case 'G': bin=4; break;
default: bin=0; break;
}
if(bin!=0) data.increment(5);
data.increment(bin);
}

/*
* print DNA histogram with data pointed to by <data>
* to stream pointed to by ptr
*/
dna_histogram(seq1 *data, const long start, const long end, FILE *ptr){
const long N=end-start+1;
long bin;
fprintf(ptr,"\n");
fprintf(ptr,"\n");
fprintf(ptr,"\n");
for(bin=1;bin<=4;bin++)fprintf(ptr,"%10.0lf\n",data->get(bin));
}

/*
* operators
*===============================================================================*/
operator dnaseq x = dnaseq y

/*
* constants
*===============================================================================*/
const long N=dnaseq();
const long M=dnaseq();
```cpp
long n;
char symbol;

if (N!=M) {
    fprintf(stderr,"\nERROR using dnaseq::operator=: length of x (%ld) differs from length of y (%ld) \n",N,M);
    exit(EXIT_FAILURE);
}
for (n=0;n<N;n++) {
    symbol = y.get(n);
    put(n,symbol);
}

/#=====================================================
extern operations
#=====================================================

/* map dna symbols to R^1
A-->1 T-->2 C-->3 G-->4 0-->0 other-->-1
*/
double dna_dnatR1(char symbol)
{
    double r;
    switch(symbol) {
      case 'A': r=1.0; break;
      case 'T': r=2.0; break;
      case 'C': r=3.0; break;
      case 'G': r=4.0; break;
      default :
        fprintf(stderr,"\nERROR using dna_dnatR1(symbol): symbol='%c' (0x%x) is not in the valid domain [A,T,C,G]\n",symbol,symbol);
        exit(EXIT_FAILURE);
    }
    return r;
}

/* map dna symbols to R^2
*/
vector dna_dnatR2(char symbol)
{
    vector r;
    switch(symbol) {
      case 'A': r.put( 1.0, 0 ); break;
      case 'T': r.put(-1.0, 0 ); break;
      case 'C': r.put( 0, +1.0 ); break;
      case 'G': r.put( 0, -1.0 ); break;
      default :
        fprintf(stderr,"\nERROR using dna_dnatR2(symbol): symbol='%c' (0x%x) is not in the valid domain [A,T,C,G]\n",symbol,symbol);
        exit(EXIT_FAILURE);
    }
    return r;
}

/* map R^2 values to dna values using Euclidean metric
*/
dnaseq dna_R2todna_euclid(seq2 xy)
{
    long n;
    int m;
    long N=xy.getN();
    double d[5];
    double smallestd;
    char closestface;
    vector2 p,q[5];
    dnaseq rdna(N);
    //q[0].put(0,0,0);
    q[1]=dna_dnatR2('A');
    q[2]=dna_dnatR2('T');
    q[3]=dna_dnatR2('C');
    q[4]=dna_dnatR2('G');
    for (n=0; n<N; n++) {
        p.put(xy.getx(n),xy.gety(n));
        smallestd=ae_metric(1,p,q[1]);
        closestface='A';
    }
}
```
for (m=2; m<5; m++) {
    d[m] = ae_metric(1, p, q[m]);
    if (((m&0x001) && (d[m]<smallestd)) || (!((m&0x001)) && (d[m]<=smallestd))) {
        // bias odd samples
        bias even samples
        // towards smaller values
        towards larger values
        smallestd=d[m];
        switch (m) {
            case 1: closestface = 'A'; break;
            case 2: closestface = 'T'; break;
            case 3: closestface = 'C'; break;
            case 4: closestface = 'G'; break;
            default: fprintf(stderr, "Error in dna_R2todna_larc(seqR2 xy)\n");
        }
    }
}

return rdna;

/*
 * map R^1 values to gsp face values using Euclidean metric
 */
dnaseq dna_R1todna_euclid(seqR1 xy) {
    long n;
    long N=xy.getN();
    char closestface;
    double p;
    dnaseq rgsp(N);
    for (n=0; n<N; n++) {
        p = xy.get(n);
        if (p<1.5) closestface = 'A';
        else if (p>3.5) closestface = 'G';
        else if (p>2.5) closestface = 'C';
        else closestface = 'T';
        rgsp.put(n, closestface);
    }
    return rgsp;
}

/*
 * dna metric d(a,b)
 */
d(a,b) | 0  A  T  C  G (b)
| 0  1  1  1  1
| A  1  0  1  1
| T  1  0  1  1
| C  1  1  0  1
| G  1  1  1  1

* On success return d(a,b). On error return -1.
* 
*/
double dna_metric(char a, char b){
    int ra=dna_dnaToRl(a);
    int rb=dna_dnaToRl(b);
    double d;
    if (ra<0) fprintf(stderr, "a=%c(0x%x) not in domain of gsp metric d(a,b)\n", a, a);
    if (rb<0) fprintf(stderr, "b=%c(0x%x) not in domain of gsp metric d(a,b)\n", b, b);
    if (ra<0) d=-1.0;
    else if (rb<0) d=-1.0;
    else if (ra==rb) d=0.0;
    else d=1.0;
    return d;
}

/*
 * real gsp metric p(x,y) where x and y are rgsp sequences computed as
 * p(x,y) = d(x0,y0) + d(x1,y1) + d(x2,y2) + ... + d(x(N-1),y(N-1))
 * where d(a,b) is defined above.
 * On success return d(x,y). On error return -1.
 */
double dna_metric(dnaseq x, dnaseq y){
    double rval, d;
    long n;
    for (m=2; m<5; m++) {
        d[m] = ae_metric(1, p, q[m]);
        if (((m&0x001) && (d[m]<smallestd)) || (!((m&0x001)) && (d[m]<=smallestd))) {
            // bias odd samples
            bias even samples
            // towards smaller values
            towards larger values
            smallestd=d[m];
            switch (m) {
                case 1: closestface = 'A'; break;
                case 2: closestface = 'T'; break;
                case 3: closestface = 'C'; break;
                case 4: closestface = 'G'; break;
                default: fprintf(stderr, "Error in dna_R2todna_larc(seqR2 xy)\n");
            }
        }
    }
    return rdna;
}

/*
 * map R^1 values to gsp face values using Euclidean metric
 */
dnaseq dna_R1todna_euclid(seqR1 xy) {
    long n;
    long N=xy.getN();
    char closestface;
    double p;
    dnaseq rgsp(N);
    for (n=0; n<N; n++) {
        p = xy.get(n);
        if (p<1.5) closestface = 'A';
        else if (p>3.5) closestface = 'G';
        else if (p>2.5) closestface = 'C';
        else closestface = 'T';
        rgsp.put(n, closestface);
    }
    return rgsp;
}

/*
 * dna metric d(a,b)
 */
d(a,b) | 0  A  T  C  G (b)
| 0  1  1  1  1
| A  1  0  1  1
| T  1  0  1  1
| C  1  1  0  1
| G  1  1  1  1

* On success return d(a,b). On error return -1.
*/
double dna_metric(char a, char b){
    int ra=dna_dnaToRl(a);
    int rb=dna_dnaToRl(b);
    double d;
    if (ra<0) fprintf(stderr, "a=%c(0x%x) not in domain of gsp metric d(a,b)\n", a, a);
    if (rb<0) fprintf(stderr, "b=%c(0x%x) not in domain of gsp metric d(a,b)\n", b, b);
    if (ra<0) d=-1.0;
    else if (rb<0) d=-1.0;
    else if (ra==rb) d=0.0;
    else d=1.0;
    return d;
}

/*
 * real gsp metric p(x,y) where x and y are rgsp sequences computed as
 * p(x,y) = d(x0,y0) + d(x1,y1) + d(x2,y2) + ... + d(x(N-1),y(N-1))
 * where d(a,b) is defined above.
 * On success return d(x,y). On error return -1.
 */
double dna_metric(dnaseq x, dnaseq y){
    double rval, d;
    long n;
long N=x.getN();
long M=y.getN();
long N= (N>M)?N:M;   /* N = the smaller of N and M */
for (n=0,d=0;n<N;++n){
    rval=dna_metric(x.get(n),y.get(n));
    if(rval<0){d+=0.0; printf("\t\%d\" ,rval);}  
    else d+=rval;
}  
if (N!=M) {
    fprintf(stderr,"ERROR using dna_metric(x,y): size of x (%ld) does not equal the size of y
\t\%d\"\n",N,M);
    exit(EXIT_FAILURE);
}  
return d;
}

/*  * autocorrelation Rxx of a dna sequence x with 2N offset */
int dna_seq::Rxx(const seqRl *rxx, const int showcount){
    long N=x.getN();
    int rval;
    rval=Rxx(rxx,showcount);
    rxx->add(2* N);
    return rval;
}

/*  * autocorrelation Rxx of a dna sequence x */
int dna_seq::Rxx(const seqRl *rxx, const int showcount){
    long m;
    const long N=x.getN();
    int rval=0;
    double rxxm;
    if(showcount)fprintf(stderr,"\t\%d auto-correlation values ... n=" 2  N+1);
    for(m= N;m<2*N;++m){
        rxxm=Rxx(m);
        if(rxxm>0) rval= 1;
        rxx->put(m,n, rxxm);
        if(showcount)fprintf(stderr, "b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b\b")
    }
    if(showcount)fprintf(stderr,"\%d .... done.\n",m N);
    return rval;
}

/*  * autocorrelation Rxx(m) */
double dna_seq::Rxx(const long m){
    const long mm= labs(m);
    const long N=x.getN();
    long n,mm;
    double d, sum;
    char a,b;
    for (n=0,sum=0; n<(N*mm) ; n++){
        mm=n-mm;
        a=(n < 0 ? n >=N)? 0.0 : get(n);
        b=(mm<0 ? mm>=N)? 0.0 : get(mm);
        d=(a==0 ? b==0)? 1.0 : dna_metric(a,b);
        sum+=d;
    }
    return sum;
}

/*  * read dna sequence from FASTA formatted file  
* and return how many symbols are in it */
* reference: https://www.genomatix.de/online_help/help/sequence_formats.html
*/

long numSym_fasta_file(const char *filename){
    FILE *fpit;
    int bufN;
    long N=0;
    char buffer[1024];
F.6  Legrange arc distance routines

#include "nseq.h"

/*====================================================================*/
/*  Legrange arc distance routines                                       */
/*  Permute a tree to minimize the distance between arcs.              */
/*  Parameter:                                                           */
/*  Current tree structure                                             */
/*  Destination tree structure                                         */
/*  Maximum number of permutations allowed                            */
/*  Wrapper function for the iteration procedure                      */
/*  Callback function to check if a permutation is allowed             */
/*  Callback function to evaluate a permutation                       */
/*====================================================================*/

#define MAX_PERMUTATIONS 100

int legrange_arc_distance(char *current_tree, char *destination_tree, int max_permutations, iteration_function, check_permutation_function, evaluate_permutation_function)
{
    int i, j, k;
    int current_energy = 0;
    int best_energy = 0;
    int best_permutation = 0;

    for (i = 0; i < max_permutations; i++)
    {
        if (check_permutation_function(current_tree, destination_tree))
        {
            current_energy = evaluate_permutation_function(current_tree, destination_tree);
            if (current_energy < best_energy)
            {
                best_energy = current_energy;
                best_permutation = i;
            }
        }
    }

    return best_energy;
}

/*====================================================================*/
/*  Iteration function for the legrange arc distance                   */
/*====================================================================*/

void iteration_function(char *current_tree, char *destination_tree)
{
    // Implementation for the iteration function
}

/*====================================================================*/
/*  Check function for permutations                                   */
/*====================================================================*/

int check_permutation_function(char *current_tree, char *destination_tree)
{
    // Implementation for the check function
}

/*====================================================================*/
/*  Evaluation function for permutations                               */
/*====================================================================*/

int evaluate_permutation_function(char *current_tree, char *destination_tree)
{
    // Implementation for the evaluation function
}

/*====================================================================*/
/*  Main function for legrange arc distance                             */
/*====================================================================*/

int main()
{
    char current_tree[100][100];
    char destination_tree[100][100];
    int max_permutations = 100;
    int best_energy = 0;

    best_energy = legrange_arc_distance(current_tree, destination_tree, max_permutations, iteration_function, check_permutation_function, evaluate_permutation_function);

    return 0;
}

/*====================================================================*/
/*  End of legrange arc distance routines                              */
/*====================================================================*/
vector2 p, q;
public:
larcc(vector2 pp, vector2 qq) {p=p, q=qq;} // constructor
larcc(double px, double py, double qx, double qy) {p.put(px,py); q.put(qx, qy);}
larcc(void) {p.put(0, 0); q.put(0, 0);}
void setp(vector2 pp) {p=pp;}
void setq(vector2 qq) {q=qq;}
void setp(double px, double py) {p.put(px, py);}
void setq(double qx, double qy) {p.put(qx, qy);}
vector2 getp(void) {return p;}
vector2 getq(void) {return q;}
double r(double theta);
vector2 x(double theta) {return r(theta)*cos(theta);}
vector2 y(double theta) {return r(theta)*sin(theta);}
vector2 xy(double theta);
double indefinite(double theta); // indefinite integral of arc length
double arclength(vector2);
double arclength(long int N);
// double operator[](double x, double y) {double z; if(x==0)return 0; else return x/y;} // division
with 0/y = 0 even when y=0
// double operator[](double x, double y) {double z; if(x==0)x=0; else z=x/y; return z;} // division
with 0/y = 0 even when y=0

/*====================================================================*/
* prototypes
*====================================================================*/
extern double larcc_arclength(double rp, double rq, double tdiff);
extern double larcc_indefinite(double rp, double rq, double thetap, double thetaq, double theta);
extern double larcc_metric(const vector2 p, const vector2 q);
extern double larcc_metric(const vector2 p, const vector2 q, const vector2 q, long int N);
extern double larcc_metric(const vector3 p, const vector3 q);
extern double larcc_metric(const vector4 p, const vector4 q);
extern double larcc_metric(const vector6 p, const vector6 q);
// extern vector2 larcc_findq(const vector2 p, const vector2 q, const vector2 theta, const double d, const double minq, const double maxq, const double maxerror, const long N);
// extern int larcc_findq(const vector2 p, const vector2 q, const vector2 theta, const double d, const double minq, const double maxq, const double maxerror, long N, vector2 *q);
extern vector3 larcc_findq(const vector3 p, const vector3 q, const vector3 theta, const double d, const double phi, const double minq, const double maxq, const double maxerror, const long N);
// extern vector3 larcc_findq(vector3 p, const vector3 theta, double phi, double d, double long int N);

/*====================================================================*/
* Daniel J. Greenhoe
* routines for Lagrange arcs
* Lagrange arcs are defined here in a manner analogous to
* Lagrange polynomial interpolation.
* Langrange polynomial interpolation is typically defined using
* Cartesian coordinates in the R^2 plane.
* Here, "Lagrange arcs" use basically the same idea, but are defined using
* polar coordinates in the R^2 plane:
* y
* | o p Let (rp, tp) be the polar location of point p.
* | / where rp is the Euclidean distance from (0,0) to p
* | / and tp is radian measure from the x-axis to p.
* /tp Let (rq, tq) be the polar location of point q.
* --------------------------- x
* \tq The "Lagrange arc" r(theta) is defined here as
* \| r(theta) = rp \----------------- + rq \----------------- + tp-tq + tq-tp
* | o q
* | theta -tq theta -tp
* |--------------------------------*(-)
*====================================================================*/
* headers
*====================================================================*/
#include<stdio.h>
#include<stdlib.h>
#include<math.h>
#include<main.h>
#include<r1.h>
#include<r2.h>
#include<r3.h>
#include<r4.h>
#include<r6.h>
#include<euclid.h>
#include<larc.h>
double larc_arclength(double rp, double rq, double tdiff)
{
  double y;
  const double phi=fabs(tdiff);
  const double rho=rq-rp;
  const double sp=sqrt(rp*rp+phi*phi-rho*rho);
  const double sq=sqrt(rp*rp+phi*phi-rho*rho);
  const double uperp=rp*phi+fabs(rho)*sp;
  const double uperq=rq*phi+fabs(rho)*sq;
  if(rp==0) {fprintf(stderr,"\nERROR using larc_arclength(rp,rq,tdiff): rp=\n\n", rp);
    exit(EXIT_FAILURE);}
  if(rq==0) {fprintf(stderr,"\nERROR using larc_arclength(rp,rq,tdiff): rq=\n\n", rq);
    exit(EXIT_FAILURE);}
  if(tdiff<=0) {fprintf(stderr,"\nERROR using larc_arclength(rp,rq,tdiff): tdiff=\n\n", tdiff);
    exit(EXIT_FAILURE);}
  if(tdiff>PI) {fprintf(stderr,"\nERROR using larc_arclength(rp,rq,tdiff): tdiff>PI\n\n", tdiff);
    exit(EXIT_FAILURE);}
  // y = ((larc_indefInt(rp,rq,0,tdiff,tdiff) - larc_indefInt(rp,rq,0,tdiff,0))/tdiff;
  // y2 = (rho>0)? (fabs(rho)/2*phi)**(log(rp*phi+sq)-log(-rho*phi+sp)):
  // (fabs(rho)/2*phi)**(log(-rho*phi+sq)-log(rp*phi+sp));
  if(fabs(rho)<.00000000001) y=rp*phi;
  else {
    if(up<0) {fprintf(stderr,"\nERROR using larc_arclength(rp,rq,tdiff): up=\n\n", up);
      exit(EXIT_FAILURE);}
    if(up>0) {fprintf(stderr,"\nERROR using larc_arclength(rp,rq,tdiff): up=\n\n", up);
      exit(EXIT_FAILURE);}
    y = (rp*sq - rp*sp)/(2*rho) + fabs(rho)*((log(sqq)-log(sq))/(2*phi));
  }
  return y;
}

double larc_indefInt(double rp, double rq, double tdiff, double thetap, double thetaq, double theta)
{
  double ra = (rp-rq);
  double rb = (rq-thetap-rp-thetaq);
  double a = rp-ra;
  double b = 2*ra*rb;
  double c = ra*ra + rb*rb;
  double x = theta;
  double y = (b+2*a*x)/(4*a)*sqrt(a*a*b*b+2*b*c)+
             (4*a*a*b*b)/(8*a*sqrt(a))+log(2*a*a*b*b+2*b*c*sqrt(a*a*b*b+2*b*c))+
             (4*a*c-b*b)/(8*a*sqrt(a)+log(fabs(2*a*a*b*b+2*b*c*sqrt(a*a*b*b+2*b*c)))); // note: fabs(...) is an error in (37)
  return y;
}

double larc_arclength(const vectR2 p, const vectR2 q)
{
  const double dp=p.q(), dq=q.q();
  const double phi = pqtheta(p.q);
  const vectR2 pq=p-q;
  double d;
  if(dp==0 || dq==0 || phi<0.0000001) return Euclidean metric
  return larc_arclength(dp, dq, phi);
}

* path length s of Lagrange arc from a point p at polar coordinate (rp,tp)
* to point q at polar coordinate (rq,tq).
* t = -------------------------
* s = | ds dtheta = | sqrt(r^2 + (----------------__) ) dtheta
* t = | tp
*/

* reference: Paul Dziewonski,
* http://tutorial.math.lamar.edu/Classes/CalcII/PolarArcLength.aspx
* https://books.google.com/books?id=b4sCQAQBAJkpg=PA533
*/

------------------------------------------------------------------------*/

double larc_arclength(double rp, double rq, double tdiff)
{
  double y;
  const double phi=fabs(tdiff);
  const double rho=rq-rp;
  const double sp=sqrt(rp*rp+phi*phi-rho*rho);
  const double sq=sqrt(rp*rp+phi*phi-rho*rho);
  const double uperp=rp*phi+fabs(rho)*sp;
  const double uperq=rq*phi+fabs(rho)*sq;
  if(rp==0) {fprintf(stderr,"\nERROR using larc_arclength(rp,rq,tdiff): rp=\n\n", rp);
    exit(EXIT_FAILURE);}
  if(rq==0) {fprintf(stderr,"\nERROR using larc_arclength(rp,rq,tdiff): rq=\n\n", rq);
    exit(EXIT_FAILURE);}
  if(tdiff<=0) {fprintf(stderr,"\nERROR using larc_arclength(rp,rq,tdiff): tdiff=\n\n", tdiff);
    exit(EXIT_FAILURE);}
  if(tdiff>PI) {fprintf(stderr,"\nERROR using larc_arclength(rp,rq,tdiff): tdiff>PI\n\n", tdiff);
    exit(EXIT_FAILURE);}
  // y = ((larc_indefInt(rp,rq,0,tdiff,tdiff) - larc_indefInt(rp,rq,0,tdiff,0))/tdiff;
  // y2 = (rho>0)? (fabs(rho)/2*phi)**(log(rp*phi+sq)-log(-rho*phi+sp)):
  // (fabs(rho)/2*phi)**(log(-rho*phi+sq)-log(rp*phi+sp));
  if(fabs(rho)<.00000000001) y=rp*phi;
  else {
    if(up<0) {fprintf(stderr,"\nERROR using larc_arclength(rp,rq,tdiff): up=\n\n", up);
      exit(EXIT_FAILURE);}
    if(up>0) {fprintf(stderr,"\nERROR using larc_arclength(rp,rq,tdiff): up=\n\n", up);
      exit(EXIT_FAILURE);}
    y = (rp*sq - rp*sp)/(2*rho) + fabs(rho)*((log(sqq)-log(sq))/(2*phi));
  }
  return y;
}

double larc_indefInt(double rp, double rq, double tdiff, double thetap, double thetaq, double theta)
{
  double ra = (rp-rq);
  double rb = (rq-thetap-rp-thetaq);
  double a = rp-ra;
  double b = 2*ra*rb;
  double c = ra*ra + rb*rb;
  double x = theta;
  double y = (b+2*a*x)/(4*a)*sqrt(a*a*b*b+2*b*c)+
             (4*a*a*b*b)/(8*a*sqrt(a))+log(2*a*a*b*b+2*b*c*sqrt(a*a*b*b+2*b*c))+
             (4*a*c-b*b)/(8*a*sqrt(a)+log(fabs(2*a*a*b*b+2*b*c*sqrt(a*a*b*b+2*b*c)))); // note: fabs(...) is an error in (37)
  return y;
}

double larc_arclength(const vectR2 p, const vectR2 q)
{
  const double dp=p.q(), dq=q.q();
  const double phi = pqtheta(p.q);
  const vectR2 pq=p-q;
  double d;
  if(dp==0 || dq==0 || phi<0.0000001) return Euclidean metric
  return larc_arclength(dp, dq, phi);
}

* path length s of Lagrange arc from a point p at polar coordinate (rp,tp)
* to point q at polar coordinate (rq,tq).
* t = -------------------------
* s = | ds dtheta = | sqrt(r^2 + (----------------__) ) dtheta
* t = | tp
*/

* reference: Paul Dziewonski,
* http://tutorial.math.lamar.edu/Classes/CalcII/PolarArcLength.aspx
* https://books.google.com/books?id=b4sCQAQBAJkpg=PA533
*/

------------------------------------------------------------------------*/
```c
    d = emetric(p,q);
    // printf("p=(%3.2lf,%3.2lf) q=(%3.1f,%3.1f) rq=%lf theta=%2.1f phi=%2.1f PI d=%lf
    aein",p.getx(),p.gety(),q.getx(),q.gety(),q.mag(),pqtheta(p,q)/PI, phi/PI,d);
    }
    else { // use Lagrange arc length
        d = larc_arclength(rp, rq, phi);
        // printf("p=(%3.2lf,%3.2lf) q=(%3.1f,%3.1f) rq=%lf theta=%2.1f phi=%2.1f
        larc\n",p.getx(),p.gety(),q.getx(),q.gety(),q.mag(),pqtheta(p,q)/PI, phi/PI,d);
    }
    return d/PI;
    }
    */

    /* Lagrange arc metric from <p> to <q> computed numerically with resolution <N>
    * Note: This function should be viewed as DEPRECATED
    * (that is, don’t use it for general computations),
    * but instead it is strongly recommended to use larc_metric(vector2 p, vector2 q).
    * The function larc_metric(vector2 p, vector2 q) uses a closed form solution
    * (from an integral lookup table).
    * This function uses a numeric estimation
    * (by an approximated summation along the arc path).
    * However, this function is still useful for testing and verification of
    * larc_metric(vector2 p, vector2 q).
    */
    double larc_metric(const vector2 p, const vector2 q, const long int N){
        larc arc(p,q);
        double d = arc.arclength(N);
        double ds=d/PI;
        return ds;
    }
}

    /* Lagrange arc metric from <p> to <q> in R^3
    */
    double larc_metric(const vector3 p, const vector3 q){
        const double rp=p.mag(), rq=q.mag();
        const double tdiff = pqtheta(p,q);
        const vector3 pq=p-q;
        double d;
        if(rp==0 || rq==0 || tdiff<=0) d = pq.mag();
        else if(rp==rq) d = rp*tdiff;
        else d = larc_arclength(rp, rq, tdiff);
        return d/PI;
    }
    */

    /* Lagrange arc metric from <p> to <q> in R^3
    */
    double larc_metric(const vector4 p, const vector4 q){
        const double rp=p.mag(), rq=q.mag();
        const double tdiff = pqtheta(p,q);
        const vector4 pq=p-q;
        double d;
        if(rp==0 || rq==0 || tdiff<=0) d = pq.mag();
        else if(rp==rq) d = rp*tdiff;
        else d = larc_arclength(rp, rq, tdiff);
        return d/PI;
    }
    /*
    */

    /* Lagrange arc metric from <p> to <q> in R^6
    */
    double larc_metric(const vector6 p, const vector6 q){
        const double rp=p.mag(), rq=q.mag();
        const double tdiff = pqtheta(p,q);
        const vector6 pq=p-q;
        double d;
        if(rp==0 || rq==0 || tdiff<=0) d = pq.mag();
        else if(rp==rq) d = rp*tdiff;
        else d = larc_arclength(rp, rq, tdiff);
        return d/PI;
    }
    */

    /* path length of arc computed using numeric integration
    */
    double larc::arclength(long int N){
`
double sum=0;
double rp=p.mag(), rq=q.mag();
double tdiff=pqtheta(p,q);
double tp=0, tq=tdiff;
double theta=tp;
long int n;
vectR2 pl,p2;
double delta=tdiff/(double)N;
vectR2 pq=p-q;
double d=pq.mag(); // Euclidean distance (p,q)
if(rp==0) return d;
if(rq==0) return d;
if(tdiff==0) return d;
for (n=0; n<N; n++){
    pl = xy(theta);
    theta += delta;
    p2 = xy(theta);
    sum += chordlength(pl,p2);
} return sum;

/*
 * find the point (x(t),y(t)) on the Lagrange arc larc(p,q) at parameter <theta>
 */
vectR2 larc::xy(double theta){
double rt=r(theta);
vectR2 pt(rt*cos(theta),rt*sin(theta));
return pt;
}

/*
 * return r(theta) for Lagrange arc (p,q)
 */
double larc::r(double theta){
double rp=p.mag();
double rq=q.mag();
double tdiff=pqtheta(p,q);
double tp=0, tq=tdiff;
double r =rp*(theta-tp)/(tp-tq) + rq*(theta-tp)/(tq-tp); // Lagrange polynomial of theta
return r;
}

/*
 * Find a point q in R^2 orientated <phi> with respect to <p>
 * that is within a <maxerror> distance <d> from the point <p>. 
 * Search for this point using <N> search locations
 * over a radial distance from <p> of <minrq> to <maxrq>. 
 * If a solution is found, place the point q at <q> and return 1.
 * If a solution is not found and an apparent discontinuity occurred in
 * the search, issue a warning and return 0.
 * If a solution is not found and a discontinuity apparently did NOT occur
 * in the search, issue an ERROR message and exit.
 */
int larc_findq(const vectR2 p, const double theta, const double d, const double minrq, const double maxrq, const double maxerror, const long N, vectR2 *q){
double rq, dd, ddprev, errord, bestrq, bestd, phi, smallesterror, discon1, discon2;
vectR2 qq, bestq;
int discontinuity=0, retval=1;
qq=polartoxy(minrq,theta);// convert polar coord. to rectangular coordinates
qq+=p;// search "origin" is the point p (not the R^2 origin (0,0))
dd=abs(larc_metric(p,qq));
smallesterror=fabs(dd-ddprev);
for(rq=minrq; rq<=maxrq; rq+=(maxrq-minrq)/(double)N){
    qq=polartoxy(rq,theta);// convert polar coord. to rectangular coordinates
    qq+=p;// search "origin" is the point p (not the R^2 origin (0,0))
    dd=larc_metric(p,qq);
    if(fabs(dd-ddprev)>(maxerror*100)){
        discontinuity=1;
        retval=0;
        discon1=ddprev;
        discon2=dd;
    }
}


```c

ddpnext = dd;
error = fabs(d-dd);
if (error < smallerror) {
    bestq = qq;
    bestq = rq;
    bestq = dd;
    smallerror = error;
    phi = ptheta(p,qq);
}

if (smallerror > maxerror) {
    if (discontinuity) {
        fprintf(stderr,": possible discontinuity,
        "
        fprintf(stderr,": possible discontinuity,
        "
        fprintf(stderr,": possible discontinuity,
        "
        fprintf(stderr,": possible discontinuity,
    } else {
        fprintf(stderr,": no apparent discontinuity but... \n"
        fprintf(stderr,": no apparent discontinuity but... \n"
        fprintf(stderr,": no apparent discontinuity but... \n"
        fprintf(stderr,": no apparent discontinuity but... \n    }
    exit(EXIT_FAILURE);
}

* Find the polar length of a point q with radial measure tq that is a
* distance <d> from the point <pq> with polar coordinates (rp, tp)
* using search resolution <N>

veet3 larc_find9(const veet3 p, const double theta, const double phi, const double d, const double

  minq, const double maxq, const double maxerror, const long int N){

    double rq, dd, error, bestq;
    veet3 bestq(0,0,0);
    veet3 q(0,0,0);
    double smallerror=10000;

    for (rq=minq; rq<=maxq; rq+=(maxq-minq)/ (double)N) {
        // convert polar coor. to rectangular coordinates
        q.polartoyxyz(rq,theta,phi);
        double p = p; // search "origin" is the point p (not the R^3 origin (0,0,0))
        dd = larc_metric(p,q);
        error = fabs(d-dd);
        if (error < smallerror) {
            bestq = q;
            bestq = rq;
            smallerror = error;
        }
    }

    if (smallerror > maxerror) {
        fprintf(stderr,": possible discontinuity,
        "
        fprintf(stderr,": possible discontinuity,
        "
        fprintf(stderr,": possible discontinuity,
        "
        fprintf(stderr,": possible discontinuity,
    } else {
        fprintf(stderr,": no apparent discontinuity but... \n"
        fprintf(stderr,": no apparent discontinuity but... \n"
        fprintf(stderr,": no apparent discontinuity but... \n"
        fprintf(stderr,": no apparent discontinuity but... \n    }
    exit(EXIT_FAILURE);
}

return bestq;
```

Back Matter

“It appears to me that if one wants to make progress in mathematics, one should study the masters and not the pupils.”

Niels Henrik Abel (1802–1829), Norwegian mathematician

“When evening comes, I return home and go to my study. On the threshold I strip naked, taking off my muddy, sweaty workaday clothes, and put on the robes of court and palace, and in this graver dress I enter the courts of the ancients and am welcomed by them, and there I taste the food that alone is mine, and for which I was born. And there I make bold to speak to them and ask the motives of their actions, and they, in their humanity reply to me. And for the space of four hours I forget the world, remember no vexation, fear poverty no more, tremble no more at death; I pass indeed into their world.”

Niccolò Machiavelli (1469–1527), Italian political philosopher, in a 1513 letter to friend Francesco Vettori.


quote: Machiavelli (1961), page 139.  
image: http://commons.wikimedia.org/wiki/File:Santi_di_Tito_-_Niccolo_Machiavelli%27s_portrait_headcrop.jpg, public domain


2016 Jun 16 6:40AM UTC Daniel J. Greenhoe


Richard Dedekind. Was sind und was sollen die zahlen? In Robert Fricke, Emmy Noether, and /:Oystein Ore, editors, *Gesammelte mathematische Werke,* pages 335–391. Druck und Verlag von Friedr. Vieweg and Sohn Akt.-Ges., Braunschweig, 1888a. URL http://resolver.sub.uni-goettingen.de/purl/?PPN23569441X. What are and what should be numbers?


Daniel J. Greenhoe. Order and metric geometry compatible stochastic processing. *PeerJ PrePrints*, February 19 2015. doi: 10.7287/peerj.preprints.844v1. URL https://peerj.com/preprints/844/. This paper has been submitted to the journal *Stochastic Systems* (http://www.i-journals.org/isy/) on 2015 February 18 and as of 2016 June 11 (more than 15 months later) no editor decision has yet been received.


Isaac Newton. Methodus differentialis. In *Analysis Per Quantitatum Series, Fluxiones ac Differentias : cum Enumeratio Linearum Tertii Ordinis*. 1711. “A photographic reproduction of the original Latin text ...and a translation are given” in Fraser (1919), pages 9–17 (Latin) and pages 18–25 (English).


Henri Poincaré. *La Science et l'hypothèse*. 1902a. URL http://fr.wikisource.org/wiki/La_Science_et_1%27hypo%CC%88t%CC%88se. (Science and Hypothesis).


REFERENCE INDEX

Aliprantis and Burkinshaw (1998), 126, 138, 140, 149, 158, 169
Adams and Franzosa (2008), 22, 127
Ammann and Escher (2008), 114
Apostol (1975), 5, 11, 24, 128, 141
Aristotle (330BC), 77
Banach (1922), 147, 149
Banach (1932b), 149
Banach (1932a), 149
orgen, Bang-Jensen and Gutin (2007), 7
Barvinok (2002), 11, 163
Beran (1985), 10
Berberian (1961), 137, 142, 162
Bessenyei and Pales (2014), 125, 135, 136
Besso (1879), 164
Bienaymé (1840), 164
Birkhoff (1933), 12, 13
Birkhoff (1938), 12
Birkhoff (1948), 12, 14
Blackman and Tukey (1958), 25
Blackman and Tukey (1959), 25
Blumenthal (1938), 125
Blumenthal (1953), 125, 131–134
Bollobás (1999), 147, 152, 163, 164, 167, 168
Bottazzini (1986), 6
Bourbaki (1939), 6
Bourbaki (1968), 5
Brenner (1985), 164
Brown and Watson (1996), 20, 127
Bruckner et al. (1997), 125, 148
Brunschwig et al. (2003), 134
Bryant (1985), 129
Bullen (1990), 164
Bullen (2003), 164, 167
Burkill (2004), 9
Burris and Sankappanavar (1981), 12, 14
Burris and Sankappanavar (2000), 14, 16, 17
Burstein et al. (2005), 123
Busemann (1955), 153, 155
Carter (2000), 164, 167
Cauchy (1821), 24, 167
Chatterji (1967), 20, 127
Cohn (2012), 5
Cohn (2002), 11
Comtet (1966), 20, 127
Comtet (1974), 6, 7, 20, 127
Copson (1968), 21, 125, 136, 153, 155
Corazza (1999), 156–158
Cori and Lascar (2001), 23
Costa et al. (2004), 123
Crammer et al. (2007), 123
Crammer et al. (2008), 123
Cristianini and Hahn (2007), 103
Crossley (2006), 170
Curry and Feldman (2010), 8
Czerwik (1993), 136
Davey and Priestley (2002), 13
Davis (2005), 5, 130, 156, 168, 169
Dedekind (1888b), 23
Dedekind (1888a), 23
Dedekind (1900), 10, 12, 14
Descartes (1637b), 1
Descartes (1637a), 1
Deza and Deza (2006), 105, 153, 157–159
Deza and Deza (2009), 21, 156–159
Deza and Deza (2014), 105, 125, 136, 153
DiBenedetto (2002), 168
Dieudonné (1969), 126, 136, 137, 153–155
Doboš (1998), 158
Dobrowolski and Mogilski (1995), 148
Domínguez-Torres (2010), 24
Domínguez-Torres (2015), 24
Durbin (2000), 7
Duthie (1942), 11
Albers and Dyson (1994), 105
Euclid (circa 300BC), 136, 153
Euler (1783), 173
Evans et al. (1967), 20, 127
Ewen (1950), iv
Ewen (1961), iv
Fagin et al. (2003b), 136
Fagin et al. (2003a), 136
Feldman and Valdez-Flores (2010), 8, 27
Fourier (1820), 162
Fraenkel (1922), 5
Fraser (1919), 174
Fréchet (1906), 136, 153
Fréchet (1928), 136, 153
Friedman (1970), 148
Fuhrmann (2012), 7
Galleoni and Garello (2010), v, 83, 99
Galvin and Shore (1984), 125, 126
Gauss (1866), 173
Gemignani (1972), 169
GenBank (2014), 2, 45, 63
Gibbons et al. (1977), 136
Giles (1987), 20, 128, 129, 140, 141, 149, 150, 153, 155, 158
<table>
<thead>
<tr>
<th>Reference</th>
<th>Year(s)</th>
<th>Page(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walmsley</td>
<td>(1920)</td>
<td>134</td>
</tr>
<tr>
<td>Waring</td>
<td>(1779)</td>
<td>108, 173</td>
</tr>
<tr>
<td>Watson and Crick</td>
<td>(1953b)</td>
<td>2, 45, 63</td>
</tr>
<tr>
<td>Watson and Crick</td>
<td>(1953a)</td>
<td>2, 45, 63</td>
</tr>
<tr>
<td>Weyl</td>
<td>(1935)</td>
<td>168</td>
</tr>
<tr>
<td>Wiener</td>
<td>(1914)</td>
<td>5</td>
</tr>
<tr>
<td>Willard</td>
<td>(1970)</td>
<td>22</td>
</tr>
<tr>
<td>Wilson</td>
<td>(1931b)</td>
<td>125</td>
</tr>
<tr>
<td>Wilson</td>
<td>(1931a)</td>
<td>153</td>
</tr>
<tr>
<td>Wolf</td>
<td>(1998)</td>
<td>5</td>
</tr>
<tr>
<td>Xia</td>
<td>(2009)</td>
<td>136</td>
</tr>
<tr>
<td>Young</td>
<td>(1912)</td>
<td>167</td>
</tr>
<tr>
<td>Zermelo</td>
<td>(1908b)</td>
<td>5</td>
</tr>
<tr>
<td>Zermelo</td>
<td>(1908a)</td>
<td>5</td>
</tr>
</tbody>
</table>
SUBJECT INDEX

2N-offset autocorrelation, 78, 81, 86
R^1 die distance linear space, 83, 85
R^2 spinner random variable, 83, 87
R^2 spinner distance linear space, 84, 84, 85, 88
R^3 die distance linear space, 84, 84, 85
R^3 die random variable, 83, 84, 86, 91
R^3 distance linear space, 84, 86, 91
R^4 DNA random variable, 83, 99, 102, 103, 104
R^6 die distance linear space, 84, 89
R^6 die random variable, 83, 84, 89, 96–98, 102
R^6 fair die distance linear space, 84
\( \alpha \)-scaled, 157
\( \alpha \)-scaled metric, 157
\( \alpha \)-truncated, 157
\( \alpha \)-truncated metric, 158
(\(X, d\)) is not translation invariant, 118
g-transform metric, 21
\( \phi \)-mean, 163
\( \frac{1}{2} \)-scaled Euclidean metric, 120
M-offset autocorrelation, 78
\(N\)-tuple, 155
\(n\)th-moment, 27
(strictly monotonically increasing in \( \phi \)), 114
\( B_1 \mathbb{R}^X \), ii
\( T_1 \mathbb{R} \)-Gyre Project, ii
\( C_1 \mathbb{R}_1 \mathbb{R}^X \), ii
GLB, 11
LUB, 11
\( \sigma \)-algebra, 8
\( \sigma \)-inframetric inequality, 124, 136, 136
\( \sigma \)-inframetric space, 124, 136
3-tuple, 6
6 element discrete metric, 22, 23
6 element ring, 22, 23
Abel, Niels Henrik, 215
absolute value, 11, 11, 108, 109, 131, 137, 149, 154, 156
absorptive, 12, 14
accumulation point, 124, 125, 169
addition operator, 7, 7, 23, 26
additive, 8
additive identity element, 31, 32, 37, 46
additive property, 9
adherent, 169
Adobe Systems Incorporated, ii
affine, 136, 163
algebra of set, 20
alphabetic order relation, 10
AM-GM inequality, 167
anti-symmetric, 10
antitone, 16, 16
arcs, 7
Aristotle, 77
arithmetic operator, 108, 109
arithmetic center, 28, 28, 33
arithmetic mean, 124, 137
arithmetic mean geometric mean inequality, 167
associates, 147
associative, 12, 147, 162
asymmetric metric, 153
autocorrelation, 77, 78, 78, 85
Avant-Garde, ii
b-metric, 136
Bacterium DNA sequence, 81
ball, 119
base, 33, 124–128, 131–133, 139, 168, 168, 169
bijection, 48
bijective, 7, 7, 14, 17, 51
board game spinner outcome subspace, 35
Bohr, Harald, 161
bound
least upper bound, 11
supremum, 11
boundary, 169
bounded, 124, 125, 129, 141, 157
bounded metric, 158
Bourbaki notation, 5
Cantor Intersection Theorem, 124
Cantor intersection theorem, 130, 130
cardinality, 76
cardinality of \( A \), 6
carica papaya, 81
Carl Spitzweg, 215
Cartesian coordinate system, 106
Cartesian product, 6
Cauchy, 124, 126, 128, 129, 132–134, 141, 144
Cauchy condition, 128
Cauchy sequence, 128
Cauchy sequences, 128
ceiling function, 11, 11
center, 3, 8, 27, 45, 74, 76
center measures, 37
chain, 10
chain rule, 109, 110
classification, 153
closed, 126, 126, 129, 130, 168, 169–171
closed ball, 126
closed interval, 11
closed set, 124, 125, 126
closed set theorem, 171
closure, 124, 125, 129, 169, 169, 170
closure point, 169
commutative, 12, 62, 138, 142, 147
commutative ring, 11
comparable, 10, 10
complete, 128, 129, 130
complex plane, 22, 23, 23, 49, 52, 64
concave, 147, 163, 163, 164
continuity, 113, 115, 116, 124, 125, 131, 134, 137, 170, 171
continuous, 117, 124, 126, 131, 131–134, 142, 144, 145, 149, 163, 164, 169, 169–171
contradiction, 143
converge, 128
convergence, 20, 124, 125, 132, 141, 171
metric space, 140
convergent, 124, 126, 128, 129, 131–134, 141, 142, 144, 170, 171
convergent sequence, 124
converges, 170, 170
converse, 6
Convex, 11
convex, 11, 23, 118, 147, 148, 150, 151, 163, 163, 164
functional, 163
strictly, 163
convex subset, 23, 77
convolution, 23, 24
coordinatewise order relation, 10
coordinatewise order relation, 10, 23
cross-correlation, 78
decreasing, 163
definitions
R1 die distance linear space, 83
R2 spinner distance linear space, 84
R3 die distance linear space, 84
R4 die distance linear space, 84
nth-moment, 27
σ-inframetric space, 136
6 element discrete metric, 23
6 element ring, 23
accumulation point, 169
alphabetical order relation, 10
arcs, 7
arithmetic center, 28
base, 33, 168
bijective, 7
Cartesian product, 6
Cauchy sequence, 128
center, 8
chain, 10
closed ball, 126
closed interval, 11
closed set, 126
closure point, 169
complex plane, 23
coordinatewise order relation, 10
dictionary order relation, 10
directed graph, 7
distance function, 153
distance space, 125
DNA outcome subspace, 29
DNA scaffold outcome subspace, 29
empty set, 4
extended probability space, 27
extended real numbers, 5
fair die outcome subspace, 28
fully ordered set, 10
function, 6
genome outcome subspace, 29
genome scaffold outcome subspace, 29
genetic center, 28
topology, 33, 168
topology induced by (X, d), 128
linear order relation, 10
metric space, 148
metric, 153
metric geometry, 32
metric linear space, 148
metric space, 136, 153
minimal center, 28
moment, 27
n-tuple, 6
near metric space, 136
Newton polynomial, 174
normed linear space, 149
one-to-one, 7
onto, 7
open ball, 32, 126
open interval, 11
open set, 126
ordered metric space, 22
ordered pair, 5
quasi-metric space, 22
ordered set, 10
ordered triple, 6
outcome center, 28
outcome subspace, 27
outcomes variance, 28
partially ordered set, 10
point of adherence, 169
poset, 10
power distance space, 135
probability space, 8
quasi-metric space, 153
real die outcome subspace, 29
real line, 22
relation, 6
sequence, 23
set of all functions, 6
set of all relations, 6
set of complex numbers, 6
set of integers, 5
set of natural numbers, 5
set of non-negative real numbers, 5
set of outcomes, 8
set of positive real numbers, 5
set of real numbers, 5
set of whole numbers, 5
spinner outcome subspace, 29
summation, 162
surjective, 7
topological space, 32
topological space induced by (X, d), 128
topology, 33, 168
topology induced by (X, d) on X, 128
totally ordered set, 10
triple, 6
undirected graph, 7
unordered metric space, 22
unordered quasi-metric space, 22
unordered set, 10
vertices, 7
weighted die outcome subspace, 28
weighted graph, 8
weighted real die outcome subspace, 29
Descartes, René, 1
diameter, 125
discrete metric preserving function, 75
dictionary order relation, 10
die outcome subspace, 78
die sequence, 85, 86, 89, 91
dilated, 157
dilated Euclidean, 133, 144
dilated metric, 157
directed, 7, 7
directed graph, 7, 43
directed metric, 153
directed PRNG state machine, 22
directed weighted graph, 2
discontinuity, 115–117
discontinuous, 116, 117, 131, 131–134, 169, 169, 170
discrete, 132, 143, 157
Discrete Fourier Transform, 96–98
discrete Fourier transform, 26
discrete metric, 2, 23, 28–30, 45, 46, 49, 51, 63, 65, 77, 155, 155
discrete metric geometry, 3
discrete metric preserving function, 75, 158
discrete metric transform on outcome subspaces, 75
discrete topology, 32, 33, 168, 170
distance, 4, 20, 78, 82, 108, 125, 125, 126, 131–134
distance function, 2, 20, 113, 117, 131, 133, 134, 153
Distance space, 4
distance space, 3, 4, 20, 21, 118, 123, 124, 125, 125–137, 141–145
distance spaces, 124
distributes, 72, 147
distributive, 13
distributive inequalities, 13
diverge, 133, 170
divergent, 170
DNA, 29, 45
DNA mapping with extended range, 65
DNA outcome subspace, 29
DNA scaffold outcome subspace, 29
dNA to linear structures, 62
domain, 6, 24, 105
dot product, 107
downsampled by a factor of, 23
dual, 10, 22
Dyson, Freeman, 105
Ebola virus DNA sequence, 81
dge weight, 8
empty set, 4, 126
Euclidean, 131–133, 143, 144
Euclidean metric, 83–85, 87–95, 105, 108, 118–120, 156
Euclidean metric space, 144
event, 8
examples
a-scaled, 157
a-scaled metric, 157
a-truncated, 157
a-truncated metric, 158
Bacterium DNA sequence, 81
board game spinner outcome subspace, 35
bounded, 157
bounded metric, 158
Coordinatewise order relation, 10
dilated, 157
dilated metric, 157
discrete, 157
discrete metric preserving function, 158
discrete metric transform on outcome subspaces, 75
DNA, 45
DNA mapping with extended range, 65
dNA to linear structures, 62
Ebola virus DNA sequence, 81
fair die, 56, 57
fair die mapping to isomorphic structure, 51
fair die mapping with extended range, 52
fair die mappings to real line and integer line, 49
fair die sequence, 78
Fourier analysis of Ebola DNA sequence, 99
GSP to complex plane, 64
GSP with Markov model, 65
high pass filtering of weighted die sequence, 94
high pass filtering of weighted real die sequence, 91
high pass filtering of weighted spinner sequence, 92
Inverse tangent metric, 21
Lagrange arc distance balls in \( \mathbb{R}^2 \), 121
Lagrange arc distance balls in \( \mathbb{R}^3 \), 121
Lagrange arc distance in \( \mathbb{R}^2 \), 120
Lagrange arc distance in \( \mathbb{R}^3 \), 120
Lagrange arc distance versus Euclidean metric, 119
LCG mappings standard ordering, 61
LCG mappings, sequential directed graph, 62
LCG mappings, sequential ordering, 61
LCG mappings, standard ordering, 59
length 1200 non-stationary die sequence with 10Hz oscillating mean, 95
length 12000 non-stationary artificial DNA sequence with 10Hz oscillating mean, 98
length 12000 non-stationary die sequence with 100Hz oscillating mean, 97
length 12000 non-stationary die sequence with 10Hz oscillating mean, 97
Lexicographical order relation, 10
linear addition, 72
linear addition with metric transform, 76
low pass filtering of fair die sequence, 89
low pass filtering of real die sequence, 85
low pass filtering of spinner sequence, 87
pair of dice and hypothesis testing, 68
pair of dice outcome subspace, 66
pair of spinner and hypothesis testing, 71
pair of spinners, 70
Papaya DNA sequence, 81
power transform, 76, 157
power transform metric, 158
radar screen, 157
radar screen metric, 158
Random DNA sequence, 80
real die mappings, 53
real die sequence, 78
ring multiplication, 73
SARS coronavirus DNA sequence, 80
snowflake, 157
snowflake transform, 76
snowflake transform metric, 158
spinner mappings, 56
spinner outcome subspace, 57
spinner sequence, 78
statistical edge detection using Haar wavelet on non-stationary artificial DNA sequence, 102
statistical edge detection using Haar wavelet on non-stationary die sequence, 101
the standard topology on the real line, 169
The usual norm, 149
Wavelet analysis of Phage Lambda DNA sequence, 103
weighted die mappings, 54
weighted die sequence, 79, 80
weighted real die sequence, 79
weighted ring, 39
weighted ring outcome subspace, 60
weighted spinner mappings, 57
weighted spinner outcome subspace, 37, 58
weighted spinner sequence, 80, 80
expectation, 3
expected value, 29
extended probability space, 27, 27
extended set of integers, 5
extension, 82, 105, 112
fair die, 2, 29, 51, 56, 57, 75, 78
fair die mapping to isomorphic structure, 51
fair die mapping with extended range, 52
fair die mappings to real line and integer line, 49
fair die outcome space, 90
fair die outcome subspace, 28, 29, 32, 49, 51, 52, 75, 84
fair die sequence, 78
Fast Wavelet Transform, 101
field, 3, 23, 24, 78, 147
field of real numbers, 7
Filter, 85–99, 91
filter, 25, 85, 87, 89, 91, 92, 94, 101
filter bank, 101
filtered, 25
finite, 6
FIR filtering, 3
First Fundamental Theorem of Calculus, 114, 116
division function, 11, 11
FontLab Studio, ii
Fourier analysis, 3
Fourier analysis of Ebola DNA sequence, 99
Free Software Foundation, ii
Fréchet product metric, 77
full period, 59
fully ordered set, 10
function, 6, 7–9, 16–18, 21, 23, 84, 170
functions
\( \mathbb{R}^3 \) die random variable, 84
\( \mathbb{R}^4 \) DNA random variable, 99, 100, 103, 104
\( \mathbb{R}^6 \) die random variable, 84, 96–98, 102
\( \frac{1}{2} \)-scaled Euclidean metric, 120
absolute value, 11, 11, 108, 109, 131, 137, 149, 154, 156
arithmetic center, 33
arithmetic mean, 137
asymmetric metric, 153
autocorrelation, 78
cardinality, 76
cardinality of A, 6
ceiling function, 11, 11
center, 3
center measures, 37
cross-correlation, 78
dicrete metric preserving function, 75
die sequence, 85, 86, 89, 91
directed metric, 153
discrete metric, 2, 23, 28, 29, 45, 46, 49, 51, 63, 65, 77, 155
discrete metric preserving function, 75
distance, 4, 20, 78, 82, 108, 125, 125, 126, 131–134
distance function, 20, 113, 117, 131, 133, 134
distance space, 118
data weight, 8
Euclidean metric, 83–85, 87–95, 105, 108, 118–120, 156
expectation, 3
floor function, 11, 11
geometric mean, 137
great circle metric, 105
harmonic mean, 137
isometry, 84, 85
isomorphism, 14
Lagrange arc distance space, 118
length M Haar scaling sequence, 25
length M Haar wavelet sequence, 25
length M high pass Haar scaling sequence, 25
length M high pass wavelet sequence, 25
length M low pass Haar scaling sequence, 25
length M low pass wavelet sequence, 25
length M low pass rectangular sequence, 25
length M number of points, 25
length M number of rectangles, 25
length M number of sections, 25
length M quadratic sequence, 25
length M quadratic sequence, 25
length M rectangular sequence, 25
length M rectangular sequence, 25
length M rectangular sequence, 25
length M rectangular sequence, 25
length M rectangular sequence, 25
length M rectangular sequence, 25
length M rectangular sequence, 25
length M rectangular sequence, 25
length M rectangular sequence, 25
length M rectangular sequence, 25
angular sequence, 92
length 50 low pass Hann-
ing sequence, 25
length 8 Haar scaling se-
quence, 26
length 8 Haar wavelet se-
quence, 26
length 9 Haar wavelet se-
quence, 26
low pass rectangular se-
quence, 87
low pass sequence, 85,
89
maximum, 137
metric, 4, 20–22, 29, 43,
77, 108, 118, 133, 156, 161
metric function, 124,
131
metric induced by the
norm, 149
metric preserving func-
tion, 74–76, 156, 156–158
metric transform, 74
minimum, 137
modulus, 11
natural log, 108, 109
Newton polynomial, 106
norm, 118, 149
outcome center, 33, 46
outcome expected value,
46, 47, 47, 49, 55
outcome random vari-
able, 46, 46, 52, 68
outcome random vari-
ables, 59
outcome subspace se-
quence metric, 77, 77, 78
outcome subspace vari-
ance, 47
outcome variance, 28,
37, 46, 47, 49
PAM die random vari-
able, 96–98, 101
PAM DNA random vari-
able, 99, 100, 102, 104
Polar angle, 118
power distance, 135
power distance function,
135, 137, 145
Power mean, 124, 135
Power mean, 124, 135,
164, 164
power triangle function,
124, 135, 135, 136, 141, 142
power triangle inequality, 4
probability function, 3,
8, 29
pullback, 48
pushforward, 75
QPSK die random vari-
able, 96–98, 101
QPSK DNA random vari-
able, 99, 100, 102, 104
QPSK spinor random
variable, 34
quadratic mean, 137
quasi-metric, 22, 43, 54,
62, 153, 153
random variable, 46, 47,
49, 53, 54, 71, 82
real die sequence, 4
RMS, 100, 102
root mean square, 100
scaling function, 101
sequence, 3, 78, 88, 92,
125, 128–130
set function, 6
slope-intercept, 106
spherical metric, 4, 105,
106, 108, 112
spinner sequence, 4, 87,
92
square root, 108, 109
traditional die random
variable, 94
traditional expected
value, 47, 49, 51, 55
traditional random vari-
able, 1, 9, 49
traditional variance, 47,
49
triangle function, 135,
144
usual metric, 1, 22, 23,
49, 51, 65, 156
variance, 28
vertex weight, 8
wavelet, 101
weighted die sequence,
94
GenBank, 80
Generalized AM-GM
inequality, 167
generalized arithmetic
mean geometric mean
inequality, 167
Generalized associative
property, 162
genome, 2, 45, 62, 63
genome outcome subspace,
29
genome scaffold outcome
subspace, 29
Genomic Signal Processing,
2, 45, 62
gemetric center, 28
gemetric inequality, 136
gemetric mean, 124, 137
GNU Octave
clear, 178, 179, 182, 185,
204
cos, 200, 212
data, 185, 204
nolig, 183, 185, 200, 204
exit, 178–182, 186, 187,
189, 190, 193, 195, 196, 200–
205, 207, 208, 210, 212, 213
function, 210, 211
functions, 178, 181
help, 207, 208
if, 79, 80, 178–185, 188–
201, 203–213
int, 178–183, 185, 188–
201, 203–209, 211–213
log, 210
mapping, 202
output, 182
polar, 209, 210, 212, 213
print, 183, 185, 200, 204
rand, 78–80, 181, 183,
201
real, 186–188, 192, 193,
202, 206
samples, 190, 191, 196,
197, 206
sin, 209, 212
sqrt, 196, 210
test, 178
using, 178, 183, 184,
190–192, 196–199, 201, 205,
206, 209, 211–213
with, 178–180, 183, 185,
188, 194, 200, 201, 204, 207,
209, 212, 213
zero, 183, 199
Golden Hind, ii
graph, 7, 7, 8, 27, 45
great circle metric, 105
greatest lower bound, 11,
11, 12
GSP, 2, 45, 62
GSP to complex plane, 64
GSP with Markov model, 65
Gutenberg Press, ii
half-open interval, 11, 131
Hardy, G.H., 161, 173
harmonic center, 28, 37
harmonic inequality, 136
harmonic mean, 124, 137
head, 7
Heuristica, ii
high pass filtering, 106
high pass filtering of
weighted die sequence, 94
high pass filtering of
weighted real die sequence,
91
high pass filtering of
weighted spinner sequence,
92
homogeneous, 11, 118, 119,
linear congruential pseudo-random number generator, 2, 22, 23
linear interpolation, 106
linear order relation, 10, 22, 23
linear space, 3, 85, 147, 147–149, 152
linearity, 109, 110, 119
linearly ordered, 1, 17–19, 49, 51, 127
linearly ordered set, 10, 12
Liquid Crystal, ii
low pass filtering, 106
low pass filtering of fair die sequence, 89
low pass filtering of real die sequence, 85
low pass filtering of spinner sequence, 87
low pass rectangular sequence, 87
low pass sequence, 85, 89
lower bound, 11, 11
Machiaielli, Niccolò, 215
maximally likely, 59
maximin, 12
maximum, 124, 137
maximun center, 28
mean square, 124
median, 13
median inequality, 13
meet, 11
meet sub-distributive, 13
Melissococcus plutonium strain 49.3 plasmid pMP19, 81
   generated by norm, 149
   induced by norm, 149
Metric ball, 20
metric function, 124, 131
metric geometry, 1, 2, 32, 49, 63
metric induced by the norm, 149
metric linear space, 148, 148, 150, 151
metric preserving function, 74–76, 156, 156–158
Metric space, 123
metric transform, 74
metrics, 159
   o-scaled metric, 157
   o-truncated metric, 158
   bounded, 158
dilated metric, 157
discrete, 155
inverse tangent, 21
power transform metric, 158
   radar screen, 158
   usual, 156
minimal center, 28
minimal inequality, 136
minimax, 12
minimax inequality, 12, 12–14, 28
minimum, 124, 137
Minkowski’s Inequality, 163
Minkowski’s inequality, 161
Minkowski’s Inequality for sequences, 167
modular, 13
Modular inequality, 14
modular inequality, 14
modularity inequality, 13
modulus, 11
moment, 27
monotone, 124
monotone, 16, 139
monotonicity, 113, 114
multiplication operation, 8
multiplication operator, 7, 23, 26
n-tuple, 6
natural log, 108, 109
near metric space, 124, 136, 136, 137
Newton interpolation, 106–108
Newton polynomial, 106, 174
non-commutative, 62
non-linearly ordered, 49
non-monotone, 158
non-negative, 11, 117, 125, 138, 143, 153, 155, 160, 161, 164
non-symmetric, 22
non-uniform, 59
nondegenerate, 11, 21, 48, 54, 117, 125, 139, 143, 149, 152–154, 157, 159–161
nonnegative, 8
norm, 118, 149, 149, 152
   usual, 149
normalized, 8
normed linear space, 118, 148, 149, 149, 151
not Cauchy, 133
not closed, 171
not convex, 118, 119
not isomorphic, 16, 84
not open, 131, 132
not order preserving, 16, 51, 84
not translation invariant, 118
not unique, 132, 142, 170
null space, 6
one-to-one, 7, 7
one-to-one and onto, 7
onto, 7, 7
open, 124, 125, 126, 126–128, 131–133, 139, 140, 168, 168, 169, 171
open ball, 32, 32, 33, 124, 125, 126, 126–128, 131–133, 138–141, 149, 150
open balls, 132, 139
open interval, 11, 131
open set, 126, 126–128, 131, 132, 140, 168, 170, 171
operations
   2N-offset autocorrelation, 78, 81, 86
   R3 spinner random variable, 83, 87
   R3 die random variable, 83, 86, 91
   R4 DNA random variable, 83
   R4 die random variable, 83, 89
   f-mean, 163
   M-offset autocorrelation, 78
addition operator, 7, 7, 23, 26
arithmetic operator, 108, 109
arithmetic mean, 124
autocorrelation, 77, 78, 85
closure, 169
convolution, 23, 24
Discrete Fourier Transform, 96–98
discrete Fourier transform, 26
dot product, 107
down sample, 91, 94
down sampled by a factor of M, 23
expected value, 29
Fast Wavelet Transform, 101
Filter, 85–89, 91
filter, 25, 85, 87, 89, 91, 92, 94, 101
filter bank, 101
filtered, 25
FIR filtering, 3
Fourier analysis, 3
geometric mean, 124
harmonic mean, 124
high pass filtering, 106
IIR filtering, 3
Interpolation, 106
join, 11
juxtaposition, 7
Lagrange interpolation, 4, 106–108
length 16 Hanning low pass sequence, 86, 88, 89
length 16 Hanning sequence, 94, 95
length 16 high pass rectangular sequence, 94
length 16 rectangular low pass sequence, 86, 88, 89
length 16 rectangular sequence, 94
length 50 Hanning high pass filter, 91, 93
length 50 Hanning low pass sequence, 86, 88, 90
length 50 Hanning sequence, 95
length 50 high pass rectangular sequence, 91
length 50 rectangular high pass filter, 91, 93
length 50 rectangular high pass sequence, 91
length 50 Rectangular low pass sequence, 89, 90
length 50 rectangular low pass sequence, 86
length 50 rectangular sequence, 95
limit, 128, 170
linear interpolation, 106
low pass filtering, 106
maximum, 124
mean square, 124
meet, 11
minimum, 124
multiplication operation, 8
multiplication operator, 7, 23, 26
Newton interpolation, 106–108
open set, 127
PAM die random variable, 82
PAM DNA random variable, 83
pulse amplitude modulation, 83
QPSK die random variable, 83
QPSK DNA random variable, 83
QPSK spiner random variable, 83, 88
quadrature phase shift keying, 83
sampled, 101
set difference, 126
set intersection, 127
SMR, 136
square-mean-root, 136
traditional die random variable, 82, 85, 89
traditional stochastic processing, 1
Voss Spectrum, 83
wavelet analysis, 3
operator, 147
order, 82
order preserving, 14, 14, 15, 16, 16, 51, 84, 85
order relation, 2, 10, 10, 23
order relations
alphabetical, 10
coordinatewise, 10
dictionary, 10
lexicographical, 10
order topology, 22
ordered distance linear space, 3, 82
ordered distance space, 2
ordered metric space, 22, 22, 23, 48, 53, 71, 72, 136
ordered pair, 5, 5, 131
ordered quasi-metric space, 22, 22, 27, 46–48
ordered set, 3, 10, 10–12, 16, 17, 20, 22
linearly, 10
totally, 10
ordered set of partitions of an integer, 12
ordered set of real numbers, 11
ordered triple, 6
oriented triangle inequality, 153
origin, 116, 117
outcome center, 27, 28, 28, 33, 37, 39, 41, 43, 46, 75
outcome expected value, 46, 47, 47, 49, 55
outcome random variable, 46, 46, 52, 68
outcome random variables, 59
outcome subspace, 27, 27–29, 43, 45–48, 56, 59, 63, 65, 68, 71, 72, 74, 77, 78, 82
outcome subspace sequence, 78
outcome subspace sequence metric, 77, 77, 78
outcome subspace variance, 47
outcome variance, 28, 28, 37, 46, 47, 49
outcomes, 8
pair of dice, 68
pair of dice and hypothesis testing, 68
pair of dice outcome subspace, 66, 68
pair of real dice, 66, 67
pair of spinner, 71, 73
pair of spinner and hypothesis testing, 71
pair of spinners, 70, 71
PAM die random variable, 82, 96–98, 101
PAM DNA random variable, 83, 99, 102, 103
Papaya DNA sequence, 81
partial order relation, 10
partially ordered set, 10
Peano’s Axioms, 23
Poincaré, Jules Henri, 27
point of adherence, 169
polar angle, 118
polar form, 107
polynomial
Lagrange, 173
Newton, 174
poset, 10
order preserving, 14
positive, 114
positive real numbers, 135
positivity, 113, 114
power distance, 135
power distance function, 135, 137, 145
Power distance space, 4
power distance space, 4, 124, 125, 135, 137–145
power distance spaces, 124
Power mean, 124, 135
power mean, 124, 135, 164, 164, 167
Power mean metrics, 159
Power mean metrics, 28
power set, 125, 168, 169
power set of X, 6
power transform, 76, 157
power transform metric, 158
power triangle function, 124, 135, 135, 136, 141, 142
power triangle inequality, 4, 124, 155, 155–158, 141–143, 145
power triangle triangle space, 137
preorder, 10
preserves joins, 14
preserves meets, 14, 15
probabilistic metric spaces, 135
probability function, 3, 8, 8, 29
probability measure, 8
probability space, 8, 8, 9, 27, 46
Proclus, 153
product rule, 109
properties
\((X, d)\) is not translation invariant, 118
(strictly monotonically increasing in \(d\)), 114
absorptive, 12, 14
additive, 8
additive property, 9
adherent, 169
affine, 136, 163
anti-symmetric, 10
antitone, 16, 16
associative, 147
associative, 12, 147, 162
bijection, 48
bijective, 7, 14, 17, 51
bounded, 124, 125, 129, 141
Cauchy, 124, 126, 128, 129, 132–134, 141, 144
Cauchy condition, 128
closed, 126, 126, 129, 130, 168, 169–171
closed set, 125
commutative, 12, 62, 138, 142, 147
comparable, 10, 10
complete, 128, 129, 130
concave, 147, 163, 163, 164
continuity, 113, 115, 116, 124, 125, 131, 134, 137, 170, 171
continuous, 117, 124, 126, 131, 131, 134–136, 142, 144, 145, 149, 163, 164, 169, 169–171
converge, 128
cconvergence, 20, 124, 125, 132, 141, 171
convergent, 124, 126, 128, 128, 129, 131–134, 141, 142, 144, 147, 170, 171
converges, 170, 170
Convex, 11
convex, 11, 23, 118, 147, 148, 150, 151, 163, 163, 164
decreasing, 163
dilated Euclidean, 133, 144
directed, 7, 7
discontinuity, 115–117
 discontinuous, 116, 117, 131, 131–134, 169, 169, 170
discrete, 132, 143
distance space, 130
distributes, 72, 147
distributive, 13
diverge, 133, 170
divergent, 170
Euclidean, 131–133, 143, 144
extension, 82, 105, 112
finite, 6
full period, 59
greatest lower bound, 12
homogeneous, 11, 118, 119, 148, 149, 152
idempotent, 12, 13, 15
identity, 147
incomparable, 10, 64, 84
increasing, 163
injective, 7, 21, 48
into, 7
isometric, 2, 3, 20, 69, 82, 84, 85
isometry, 21
isomorphic, 2, 3, 14, 14, 16, 52, 56–59, 61, 62, 68, 71, 72, 74, 82, 85
isotone, 16, 16, 17, 20, 157
join super-distributive,
least upper bound, 12
limit, 128
linear, 12, 17, 59, 62
linearity, 109, 110, 119
linearly ordered, 1, 17–19, 49, 51, 127
maximally likely, 59
median inequality, 13
meet sub-distributive,
metric, 82
modular, 13
monotone, 124
monotone, 16, 139
monotonicity, 113, 114
non-commutative, 62
non-linearly ordered, 49
non-monotone, 158
non-negative, 11, 117, 125, 138, 143, 153, 155, 160, 161, 164
non-symmetric, 22
non-uniform, 59
nondegenerate, 11, 21, 48, 54, 117, 125, 139, 143, 149, 153, 154, 157, 159–161
nonnegative, 8
normalized, 8
not Cauchy, 133
not closed, 171
not convex, 118, 119
not isomorphic, 16, 84
not open, 131, 132
not order preserving, 16, 51, 84
not translation invariant, 118
not unique, 132, 142, 170
one-to-one, 7
one-to-one and onto, 7
onto, 7
open, 124, 126, 126–128, 131–133, 139, 140, 168, 168, 169
open set, 128
order, 82
order preserving, 14, 14, 15, 16, 16, 51, 84, 85
oriented triangle inequality, 153
positive, 114
positivity, 113, 114
power triangle inequality, 137, 138, 145
preserves joins, 14
preserves meets, 14, 15
pseudo-distributes, 147
randomness, 8
real, 149
reflexive, 10
strictly antitone, 16, 16,
18, 32, 36, 38, 40, 42, 44
strictly convex, 163
strictly isotone, 16, 16–20, 31, 35, 36, 39–44, 46, 48, 50, 52, 54, 57, 60, 62, 64, 65, 74
strictly monotone, 135
strictly monotone, 16,
129, 136, 138, 141, 144, 164
strictly monotonic, 163
strictly monotonically increasing, 111, 114, 241, 249
strictly positive, 149
subadditive, 11, 21, 148,
149, 153–157, 159, 161
subadditive, 160, 161
submultiplicative, 11
surjective, 7, 59
symmetric, 9, 47, 57, 58,
64, 71, 113, 117, 125, 138, 142,
150, 153
symmetry, 21, 113, 155,
159, 161
transitive, 10
translation invariant,
118, 148, 151, 152
triangle inequality, 4, 11,
92, 118, 123–125, 131–134,
142, 145, 153–155, 160
triangle inequality fails,
118
triangle inequality, 149
uncorrelated, 78, 80, 91
undirected, 7, 7
uniform, 59
uniformly distributed, 78, 80
unique, 124, 126, 129–131, 133, 134, 144, 170
unordered, 23, 28, 29, 49, 84
pseudo-distributes, 147
pseudo-random number generator, 59
pstricks, ii
pullback, 48
pullback function, 75
Pullback metric, 21
pulse amplitude modulation, 83
pushforward, 48
QPSK die random variable, 83, 96–98, 101
QPSK DNA random variable, 83, 99, 100, 102, 104
QPSK spinner random variable, 83, 84, 88
quadratic inequality, 136
quadratic mean, 137
quadrature phase shift keying, 83
quasi-metric, 22, 43, 54, 62, 153, 153
quasi-metric space, 22, 153
quotes
Abel, Niels Henrik, 215
Aristotle, 77
Bohr, Harald, 161
Descartes, René, 1
Dyson, Freeman, 105
Hardy, G.H., 161, 173
Housman, Alfred Edward, iii
Jacobi, Carl Gustav Jacob, 147
Lagrange, Joseph Louis, 85
Langrange, Joseph Louis, 123
Machiavelli, Niccolò, 215
Poincaré, Jules Henri, 27
Proclus, 153
Russell, Bertrand, iii
Stravinsky, Igor, iii
Weyl, Hermann, 168
quotient structures, 11
radar screen, 157
radar screen metric, 158
Random DNA sequence, 80
random variable, 9, 46–49, 53, 54, 71, 72, 82
randomness, 8
range, 6, 80, 156
real, 149
real dice, 68
real die, 3, 4, 29, 67, 75, 78, 79, 82
real die mappings, 53
real die outcome subspace, 29, 31, 32, 53, 75, 82, 84
real die sequence, 4, 78, 106
real line, 1–3, 22, 22, 23, 46, 47, 49–51, 54, 57–59, 63, 64, 68, 71, 72, 82, 127
real line ordered metric space, 48
real numbers, 149
real part, 23
reflexive, 10
relation, 6, 6, 7, 84, 135
inverse, 6
relations
σ-infarmetic inequality, 124, 136, 136
converse, 6
coordinatewise order relation, 23
distance function, 2
domain, 6
dual, 10, 22
geometric inequality, 136
harmonic inequality, 136
image set, 6
infarmetic inequality, 136, 136
inverse, 6
lexicographical order relation, 23
minimal inequality, 136
null space, 6
order relation, 2, 10, 23
partial order relation, 10
power triangle inequality, 124, 135, 135, 136, 141, 142, 145
preorder, 10
quadratic inequality, 136
range, 6
relaxed triangle inequality, 124, 136, 136
square mean root inequality, 136
standard ordering relation, 83
triangle inequality, 123, 135, 136, 136
triangle relation, 143, 144
relaxed triangle inequality, 124, 136, 136
Reproducible Research, 177
ring, 156
ring multiplication, 73
RMS, 100, 102
root mean square, 100
Russell, Bertrand, iii
Saint Andrew’s University’s, 177
sampled, 101
SARS coronavirus DNA sequence, 80
scaffold DNA, 29
scaling function, 101
Second Fundamental Theorem of Calculus, 114
sequence, 3, 23, 23–26, 77, 78, 88, 92, 125, 128–130, 149
sequences
Cauchy, 128
sequential characterization, 131
sequential characterization of continuity, 131, 171
set, 4–8, 11, 126, 153
set difference, 126
set function, 6
set intersection, 127
set of all functions, 6
set of all relations, 6
set of complex numbers, 6, 23
set of integers, 5, 23
set of natural numbers, 5
set of non-negative real numbers, 5
set of outcomes, 8
set of positive real numbers, 5
set of real numbers, 1, 5, 22
set of whole numbers, 5
sets, 6
open ball, 149
real numbers, 149
sigma-algebra, 20
slope-intercept, 106
SMR, 136
snowflake, 157
snowflake transform, 76
snowflake transform metric, 158
source code, 177
space, 108
linear, 147
metric, 126, 149, 153, 154
metric vector, 148
normed vector, 149
topological, 168
vector, 147
spherical metric, 4, 105, 106, 108, 112
spinner, 29, 69, 70, 75, 78
spinner mappings, 56
spinner outcome subspace, 29, 35, 57, 75, 76, 84, 88
spinner sequence, 4, 78, 87, 92, 93, 106
spinners, 71
square mean root inequality, 136
square root, 108, 109
square-mean-root, 136
standard ordered set of real numbers, 11
standard ordering relation, 83
statistical edge detection using Haar wavelet on non-stationary artificial DNA sequence, 102
statistical edge detection using Haar wavelet on non-stationary die sequence, 101
Stravinsky, Igor, III
strictly antitone, 16, 16, 18, 32, 36, 38, 40, 42, 44
strictly convex, 163, 163
strictly isotone, 16, 16–20, 31, 35, 36–44, 46, 48, 50, 52, 54, 57, 60, 62, 64, 65, 74
strictly monotonically increasing, 111, 114, 241, 249
structures
\( \mathbb{R}^1 \) die distance linear space, 83, 85
\( \mathbb{R}^2 \) spinner distance linear space, 84, 84, 85, 88
\( \mathbb{R}^3 \) die distance linear space, 84, 84, 85
\( \mathbb{R}^5 \) distance linear space, 86, 91
\( \mathbb{R}^6 \) die distance linear space, 84, 89
\( \mathbb{R}^6 \) fair die distance linear space, 84
g-Transform metric, 21
N-tuple, 155
\( \alpha \)-moment, 27
\( \sigma \)-algebra, 8
\( \sigma \)-inframetric space, 124, 136
3-tuple, 6
6 element discrete metric, 22, 23
6 element ring, 22, 23
accumulation point, 124, 125, 169
algebra of set, 20
arcs, 7
arithmetic center, 28, 28
b-metric, 136
ball, 119
base, 33, 124–128, 131–133, 139, 168, 168, 169
boundary, 169
Cartesian coordinate system, 106
Cartesian product, 6
Cauchy sequence, 128
center, 8, 27, 45, 74, 76
chain, 10
closed ball, 126
closed interval, 11
closed set, 124, 126
closure, 124, 125, 129,
169, 170
closure point, 169
commutative ring, 11
complex plane, 22, 23, 23, 49, 52, 64
convergent sequence, 124
covex subset, 23, 77
die outcome subspace, 78
die sequence, 91
directed graph, 7, 43
directed PRNG state machine, 22
directed weighted graph, 2
discrete metric, 30
discrete metric geometry, 3
discrete topology, 32, 33, 168, 170
distance function, 131
Distance space, 4
distance space, 3, 4, 20, 21, 123, 124, 125, 125–137, 141–145
distance spaces, 124
DNA, 29
DNA outcome subspace, 29
DNA scaffold outcome subspace, 29
domain, 24, 105
empty set, 4, 126
Euclidean metric space, 144
extended probability space, 27, 27
extended real numbers, 5, 135
extended set of integers, 5
extension, 105
fair die, 2, 29, 51, 75, 78
fair die outcome subspace, 90
fair die outcome subspace, 28, 29, 32, 49, 51, 52, 75, 84
field, 3, 23, 24, 78, 147
field of real numbers, 7
fully ordered set, 10
function, 6, 7–9, 16–18,
21, 23, 84, 170
genome, 2, 45, 62, 63
genome outcome subspace, 29
genome scaffold outcome subspace, 29
Genomic Signal Processing, 2, 45, 62
geometric center, 28
graph, 7, 7, 8, 27, 45
GSP, 2, 45, 62
half-open interval, 11, 131
harmonic center, 28, 37
horizontal half-open interval, 131, 143
identity, 147
indiscrete topological space, 128
indiscrete topology, 168, 170
inframetric space, 136, 137
integer line, 22, 23, 23, 49, 50, 57, 60, 63, 64, 82
interior, 124, 125, 169
169
interior point, 169
interval topology, 22
inverse, 147
Lagrange arc distance space, 118
lattice, 12, 12–14, 17–19
least upper bound, 11
limit, 170
linear congruential, 59
linear congruential pseudo-random number generator, 2, 22, 23
linear order relation, 22, 23
linear space, 3, 85, 147–149, 152
linearly ordered set, 10, 12
maxmin center, 28
metric, 21, 22, 28, 32
148, 156, 159
Metric ball, 20
subject index

subspace, 75
  weighted die, 3, 29, 55, 75, 79, 80
  weighted die outcome
  subspace, 28, 29, 54, 76
  weighted graph, 3, 8, 8, 29, 31, 82
  weighted graphs, 82
  weighted real die, 29, 79, 80
  weighted real die
  outcome subspace, 29, 29, 33, 37
  weighted real die sequence, 106
  weighted spinner, 73
  weighted spinner outcome
  subspace, 37, 57
  weighted spinner sequence, 106
subadditive, 11, 21, 148, 149, 153, 155–157, 159, 161
subadditive, 160, 161
submultiplicative, 11
summation, 162
surface of a sphere with radius r, 105
surface of a sphere..., 105
surjective, 7, 7, 59
symmetric, 9, 47, 57, 58, 64, 71, 113, 117, 125, 138, 142, 150, 153
symmetry, 21, 113, 155, 159, 161
tail, 7
  The Archimedean Property, 138
  The Book Worm, 215
  The Closed Set Theorem, 170
  The principle of Archimedes, 138
  the standard topology on the real line, 169
  The usual norm, 149
  theorems
  g-transform metric, 21
  Cantor intersection, 130
  Cantor Intersection
  Theorem, 124
  Cantor intersection theo-
  rem, 130
  chain rule, 109, 110
  closed set theorem, 171
distributive inequalities, 13
  First Fundamental The-
  orem of Calculus, 114, 116
  Fréchet product metric, 77
  Generalized AM-GM in-
 equality, 167
  Generalized associative
property, 162
  inverse image character-
  ization, 131
  inverse image characterization of continuity, 131, 171
  Jensen’s Inequality, 163,
  167
  L’Hôpital’s rule, 116,
  117
  l’Hôpital’s rule, 166
  minimax inequality, 12, 12–14
  Minkowski’s Inequality,
  163
  Minkowski’s inequality,
  161
  Minkowski’s inequality for sequences, 167
  Modular inequality, 14
  Power mean metrics,
  159
  product rule, 109
  Pullback metric, 21
  Second Fundamental
  Theorem of Calculus, 114
  sequential characterization, 131
  sequential characterization of continuity, 131, 171
  The Archimedean Property,
  138
  The Closed Set The-
  orem, 170
  The principle of Archimedes, 138
  Uniqueness of limit, 142
  Young’s Inequality, 167
topological space, 32, 124,
  125, 127, 131, 142, 168–171
topological space (X, T) in-
  duced by (R, d), 132
topological space induced by
  (R, d), 131
topological space induced by
  (X, d), 128, 128, 129, 140
topologies
  discrete, 168
  indiscrete, 168
  trivial, 168
topological space, 127
topology, 20, 33, 124, 125,
  127, 128, 131–133, 168,
  168
  topology induced by (X, d)
on X, 128
totally ordered set, 10
traditional die random vari-
  able, 82, 85, 89, 94
traditional expected value, 9,
  9, 47, 49, 51, 55
traditional random variable,
  1, 9, 9, 49
  traditional stochastic pro-
  cessing, 1
  traditional stochastic process-
  ing, 1
  traditional stochastic process-
  ing, 1
  traditional stochastic process-
  ing, 1
  transition, 10
  translation invariant, 118, 148, 151, 152
  triangle function, 135, 144
  triangle inequality, 4, 11, 92,
  118, 123–125, 131–135, 136,
  136, 142, 145, 149, 153–155, 160
triangle inequality fails, 118
triangle inequality, 149
  triangle relation, 135, 143,
  144
  triangle-inequality, 152
  triple, 6
  trivial topology, 168
uncorrelated, 78, 80, 91
undirected, 7, 7
undirected graph, 7, 43
uniform, 59
uniformly distributed, 78, 80
unique, 124, 126, 129–131,
133, 134, 144, 170
Uniqueness of limit, 142
unit ball, 119
unit Lagrange arc distance
ball, 122
unordered, 23, 28, 29, 49, 84
unordered metric space, 22,
22, 23, 51, 52
unordered quasi-metric space,
22, 22, 23
unordered set, 10
upper bound, 11, 11
usual metric, 1, 21–23, 49, 51,
65, 156, 156
usual metric space, 169
usual norm, 149
Utopia, ii
values
  GLB, 11
  LUB, 11
  additive identity element,
  31, 32, 37, 46
  contradiction, 143
diameter, 125
greatest lower bound, 11
head, 7
imaginary part, 23
least upper bound, 11
limit, 129
limit point, 170
lower bound, 11, 11
outcome expected value, 47
real part, 23
tail, 7
  traditional expected value, 9, 9, 47
  traditional variance, 9, 9
  upper bound, 11, 11
variance, 28
vertical half-open interval, 131, 143
vertice weight, 8
vertices, 7
Voss Spectrum, 83
wagon wheel, 75
wagon wheel outcome subspace, 75
wagon wheel output subspace, 75
wavelet, 101

wavelet analysis, 3
Wavelet analysis of Phage Lambda DNA sequence, 103
weighted, 164
weighted die, 3, 29, 55, 75, 79, 80
weighted die mappings, 54
weighted die outcome subspace, 28, 28, 54, 76
weighted die sequence, 79, 80, 94
weighted graph, 3, 8, 8, 29, 31, 82
weighted graphs, 82
weighted real die, 29, 79, 80
weighted real die outcome subspace, 29, 29, 33, 37
weighted real die sequence, 79, 106
weighted ring, 39
weighted ring outcome subspace, 60
weighted spinner, 73
weighted spinner mappings, 57
weighted spinner outcome subspace, 37, 37, 57, 58
weighted spinner sequence, 80, 80, 106
Weyl, Hermann, 168
Who Was There, 177
Wáng Hán Zōng Zhōng Mǐng Ti Fán, ii
Young's Inequality, 167