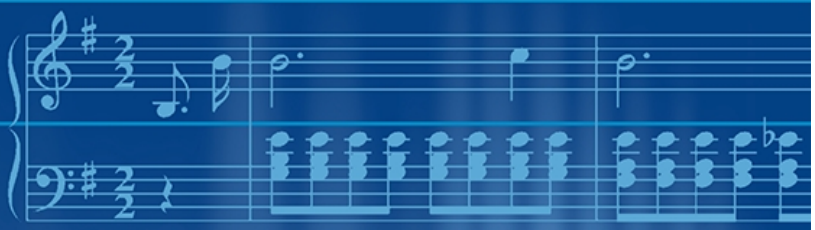


Meinard Müller

Fundamentals of Music Processing

Audio, Analysis, Algorithms, Applications

Exercises and Solutions

Springer

Meinard Müller

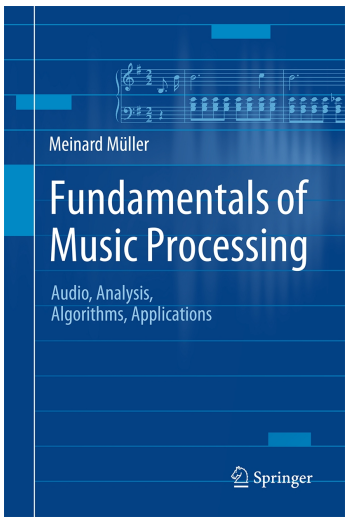
Fundamentals of Music Processing

Audio, Analysis, Algorithms, Applications

Exercises and Solutions

Springer

This manuscript contains a complete set of solutions to all exercises contained in the following textbook:



Meinard Müller
Fundamentals of Music Processing
Audio, Analysis, Algorithms, Applications
483 p., 249 illus., 30 illus. in color, hardcover
ISBN: 978-3-319-21944-8
ISBN 978-3-319-21945-5 (eBook)
DOI 10.1007/978-3-319-21945-5
Springer, 2015

Author:

Meinard Müller is professor for Semantic Audio Processing at the International Audio Laboratories Erlangen, Germany, a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and the Fraunhofer Institute for Integrated Circuits IIS. His research interests include music processing, music information retrieval, audio signal processing, multimedia processing, and motion retrieval.

Contact:

Meinard Müller
Friedrich-Alexander Universität Erlangen-Nürnberg
International Audio Laboratories Erlangen
Lehrstuhl Semantic Audio Processing
Am Wolfsmantel 33, 91058 Erlangen
meinard.mueller@audiolabs-erlangen.de

Website:

www.music-processing.de

Preface

This is the solutions manual for the textbook *Fundamentals of Music Processing* published by Springer in 2015. It contains solutions to all exercises contained in the book. This release was created on July 21, 2015. Future releases with corrections to errors will be published on the book's website (see below).

I would like to thank the various people who have provided valuable feedback on this document (and earlier releases) including the students from my courses held at the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU). I welcome all comments, questions, and suggestions about the solutions as well as reports on (potential) errors in the text or equations in this document; please send any such feedback to

meinard.mueller@audiolabs-erlangen.de

Further information and additional material for the book is available from

www.music-processing.de

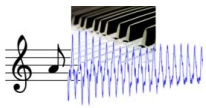
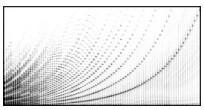
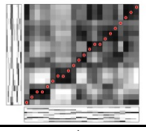
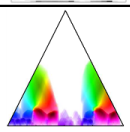
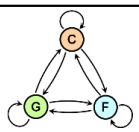
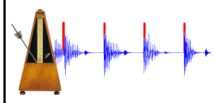
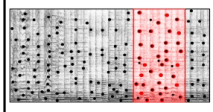
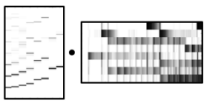
Erlangen, *Meinard Müller*
July 2015

Overview

This textbook provides both profound technological knowledge and a comprehensive treatment of essential topics in music processing and music information retrieval. Including numerous examples, figures, and exercises, this book is suited for students, lecturers, and researchers working in audio engineering, computer science, multimedia, and musicology.

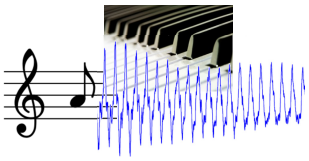
The book consists of eight chapters. The first two cover foundations of music representations and the Fourier transform—concepts that are then used throughout the book. In the subsequent chapters, concrete music processing tasks serve as a starting point. Each of these chapters is organized in a similar fashion and starts with a general description of the music processing scenario at hand before integrating it into a wider context. It then discusses, in a mathematically rigorous way, important techniques and algorithms that are generally applicable to a wide range of analysis, classification, and retrieval problems. At the same time, the techniques are directly applied to a concrete music processing task. By mixing theory and practice, the book's goal is to convey both profound technological knowledge and a solid understanding of music processing applications. Each chapter ends with a section that includes links to the research literature, suggestions for further reading, a list of references, and exercises. The chapters are organized in a modular fashion, thus offering lecturers and readers many ways to choose, rearrange or supplement the material. Accordingly, selected chapters or individual sections can easily be integrated into courses on general multimedia, information science, signal processing, music informatics, or the digital humanities.

The following figure gives an overview of the individual chapters and the main topics.

Chapter		Music Processing Scenario	Notions, Techniques & Algorithms
1		**Music Representations**	Music notation, MIDI, audio signal, waveform, pitch, loudness, timbre
2		**Fourier Analysis of Signals**	Discrete/analog signal, sinusoid, exponential, Fourier transform, Fourier representation, DFT, FFT, STFT
3		**Music Synchronization**	Chroma feature, dyamic programming, dyamic time warping (DTW), alignment, user interface
4		**Music Structure Analysis**	Similarity matrix, repetition, thumbnail, homogeneity, novelty, evaluation, precision, recall, F-measure, visualization, scape plot
5		**Chord Recognition**	Harmony, music theory, chords, scales, templates, hidden Markov model (HMM), evaluation
6		**Tempo and Beat Tracking**	Onset, novelty, tempo, tempogram, beat, periodicity, Fourier analysis, autocorrelation
7		**Content-Based Audio Retrieval**	Identification, fingerprint, indexing, inverted list, matching, version, cover song
8		**Musically Informed Audio Decomposition**	Harmonic/percussive component, signal reconstruction, instanteneous frequency, fundamental frequency (F0), trajectory, nonnegative matrix factorization (NMF)

Contents

Chapter 1
Music Representations

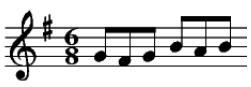


Exercise 1.1. Assume that a pianist exactly follows the specifications given in the Beethoven example from Figure 1.1. Determine the duration (in milliseconds) of a quarter note and a measure, respectively.

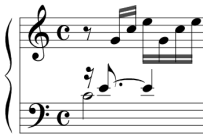
Solution to Exercise 1.1. The tempo is given by the metronome specification of 108 half notes per minute. Therefore, a measure (which equals a half note) has a duration of $1000 \cdot 60/108 = 555.56$ ms. Furthermore, a quarter note has a duration of 277.78 ms.

Exercise 1.2. Specify the MIDI representation (in tabular form) and sketch the piano-roll representation (similar to Figure 1.13) of the following sheet music representations. Assume that a quarter note corresponds to 120 ticks. Set the velocity to a value of 100 for all active note events. Furthermore, assign the notes of the G-clef to channel 1 and the notes of the F-clef to channel 2.

(a)



(b)

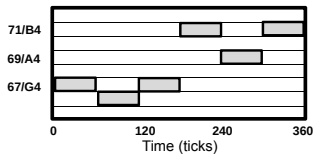


[**Hint:** In this exercise, we assume that the reader has some basic knowledge of Western music notation.]

Solution to Exercise 1.2.

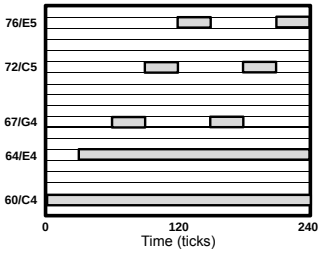
(a)

Time (Ticks)	Message	Channel	Note Number	Velocity
0	NOTE ON	1	67	100
60	NOTE OFF	1	67	0
0	NOTE ON	1	66	100
60	NOTE OFF	1	66	0
0	NOTE ON	1	67	100
60	NOTE OFF	1	67	0
0	NOTE ON	1	71	100
60	NOTE OFF	1	71	0
0	NOTE ON	1	69	100
60	NOTE OFF	1	69	0
0	NOTE ON	1	71	100
60	NOTE OFF	1	71	0

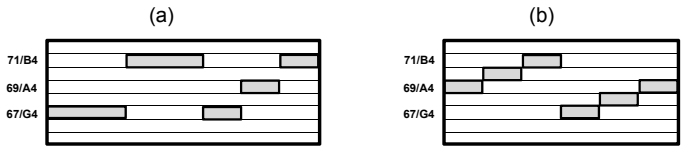


(b)

Time (Ticks)	Message	Channel	Note Number	Velocity
0	NOTE ON	2	60	100
30	NOTE ON	2	64	100
30	NOTE ON	1	67	100
30	NOTE OFF	1	67	0
0	NOTE ON	1	72	100
30	NOTE OFF	1	72	0
0	NOTE ON	1	76	100
30	NOTE OFF	1	76	0
0	NOTE ON	1	67	100
30	NOTE OFF	1	67	0
0	NOTE ON	1	72	100
30	NOTE OFF	1	72	0
0	NOTE ON	1	76	100
30	NOTE OFF	1	76	0
0	NOTE OFF	2	64	0
0	NOTE OFF	2	60	0

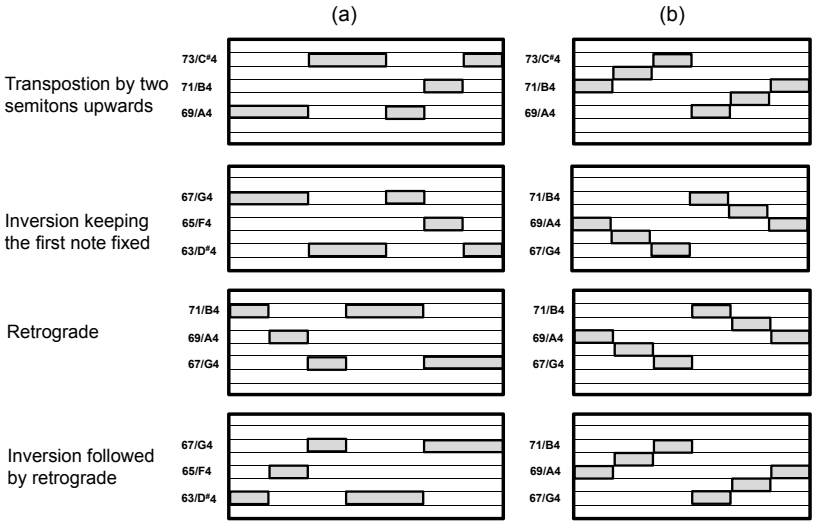


Exercise 1.3. In this exercise, a **melody** is regarded as a linear succession of musical notes. A **transposition** of a given melody moves all notes up or down in pitch by a constant interval. Furthermore, an **inversion** of a melody turns all the intervals upside-down. For instance, if the original melody rises by three semitones, the inverted melody falls by three semitones. Finally, the **retrograde** of a melody is the reverse, where the notes are played from back to front. Let us consider the following two melodies given in piano-roll representation:



Specify for each of the two melodies the piano-roll representation of the transposition by two semitones upwards, the inversion (keeping the first note fixed), the retrograde, and the retrograde of the inversion. Furthermore, regarding melodies only up to pitch classes (by ignoring octave information), determine the number of different melodies that can be generated by successively applying an arbitrary number of transpositions, inversions, and retrogrades.

Solution to Exercise 1.3.



For the fist melody, one can generate 48 different melodies (ignoring octave information). For the second melody, inversion and retrograde lead to the same melody. Altogether, one obtains 24 different melodies (ignoring octave information).

Exercise 1.4. The **speed of sound** is the distance traveled per unit of time by a sound wave propagating through an elastic medium. Look up the speed of sound in air. Assume that a concert hall has a length of 50 meters. How long does it take for a sound wave to travel from the front to the back of the hall?

Solution to Exercise 1.4. In dry air at 20°C (68°F), the speed of sound is 343.2 meters per second. For a distance of 50 meters, a sound wave requires roughly 145.7 ms.

Exercise 1.5. Using (1.1), compute the center frequencies for all notes of the C-major scale C4, D4, E4, F4, G4, A4, B4, C5 and for all notes of the C-minor scale C4, D4, E♭4, F4, G4, A♭4, B♭4, C5 (see also Figure 1.5).

Solution to Exercise 1.5.

C-major scale		
Note	p	$F_{\text{pitch}}(p)$
C4	60	261.63
D4	62	293.66
E4	64	329.63
F4	65	349.23
G4	67	392.00
A4	69	440.00
B4	71	493.88
C5	72	523.25

C-minor scale		
Note	p	$F_{\text{pitch}}(p)$
C4	60	261.63
D4	62	293.66
E♭4	63	311.13
F4	65	349.23
G4	67	392.00
A♭4	68	415.30
B♭4	70	466.16
C5	72	523.25

Exercise 1.6. Using (1.1), compute the frequency ratio $F_{\text{pitch}}(p+1)/F_{\text{pitch}}(p)$ of two subsequent pitches $p+1$ and p (see (1.2)). How does the frequency $F_{\text{pitch}}(p+k)$ for some $k \in \mathbb{Z}$ relate to $F_{\text{pitch}}(p)$? Furthermore, derive a formula for the distance (in semitones) for two arbitrary frequencies ω_1 and ω_2.

Solution to Exercise 1.6. The ratio is computed via

$$\begin{aligned} F_{\text{pitch}}(p+1)/F_{\text{pitch}}(p) &= 2^{(p+1-69)/12} \cdot 440 \cdot 2^{-(p-69))/12} \cdot (1/440) \\ &= 2^{1/12} \cdot 2^{(p-69)/12} \cdot 2^{-(p-69))/12} \\ &= 2^{1/12} \approx 1.059463. \end{aligned}$$

Futhermore, one obtains

$$F_{\text{pitch}}(p+k) = 2^{k/12} \cdot F_{\text{pitch}}(p).$$

As in (1.4), the distance (in semitones) between two frequencies ω_1 and ω_2 is

$$\log_2\left(\frac{\omega_1}{\omega_2}\right) \cdot 12.$$

Exercise 1.7. Let us have a look at Figure 1.18b, which shows a waveform obtained from a recording of Beethoven's Fifth. Estimate the fundamental frequency of the sound played by counting the number of oscillation cycles in the section between 7.3 and 7.8 seconds. Furthermore, determine the musical note that has a center frequency closest to the estimated fundamental frequency. Compare the result with the sheet music representation of Figure 1.1.

Solution to Exercise 1.7. The section between 7.3 and 7.8 seconds contains roughly 37 oscillation cycles. This corresponds to a fundamental frequency of 74 Hz. This frequency is closest to the musical note D2 ($p = 38$), which has a center frequency of $F_{\text{pitch}}(38) = 73.4$ Hz. This is the lowest note of the fourth and fifth measure shown in Figure 1.1.

Exercise 1.8. Assume an equal-tempered scale that consists of 17 tones per octave and a reference pitch $p = 100$ having a center frequency of 1000 Hz. Specify a formula similar to (1.1), which yields the center frequencies for the pitches $p \in [0:255]$. In particular, determine the center frequency for the pitches $p = 83$, $p = 66$, and $p = 49$ in this scale. What is the difference (in cents) between two subsequent pitches in this scale?

Solution to Exercise 1.8. As in (1.1), one obtains

$$F^{17}_{\text{pitch}}(p) = 2^{(p-100)/17} \cdot 1000.$$

In particular, one has $F^{17}_{\text{pitch}}(83) = 500$, $F^{17}_{\text{pitch}}(66) = 250$, and $F^{17}_{\text{pitch}}(59) = 125$. By (1.4), the difference (in cents) between two subsequent pitches is given by

$$\log_2\left(\frac{F^{17}_{\text{pitch}}(p+1)}{F^{17}_{\text{pitch}}(p)}\right) \cdot 1200 = \log_2(2^{1/17}) \cdot 1200 = 1200/17 \approx 70.6.$$

Exercise 1.9. Write a small computer program to calculate the differences (in cents) between the first 16 harmonics of the note C2 and the center frequencies of the closest notes of the twelve-tone equal-tempered scale (see Figure 1.20). What are the corresponding differences when considering the harmonics of another note such as B♭4?

Solution to Exercise 1.9. For some pitch p, the center frequency of the m^{th} harmonic, $m \in \mathbb{N}_0$, is given by $m \cdot F_{\text{pitch}}(p)$. Furthermore, by Exercise 1.6, the center frequency of some pitch $p+k$, $k \in \mathbb{Z}$, is given by $F_{\text{pitch}}(p+k) = 2^{k/12} \cdot F_{\text{pitch}}(p)$. Therefore, by (1.4), the difference (in cents) between the m^{th} harmonic of pitch p and the closest note of the twelve-tone equal-tempered scale is given by

$$\min_{k\in\mathbb{Z}}\left(\log_2\left(\frac{m \cdot F_{\text{pitch}}(p)}{2^{k/12} \cdot F_{\text{pitch}}(p)}\right) \cdot 1200\right) = \min_{k\in\mathbb{Z}}\left(\log_2(m) - \frac{k}{12}\right) \cdot 1200.$$

This shows that the differences are independent of the pitch p. In other words, the differences are the same when starting with the note C2 or when starting with another note such as B♭4. The differences can be computed by the following computer program using MATLAB:

```
for m=1:16
    diff = (log2(m)-1/12)*1200;
```

```
    diff = rem(diff,100);
    if diff > 50
        diff = diff-100;
    end
    fprintf('m = %2i, diff = %+6.2f \n ',m,diff);
end
```

This program yields the following output:

```
m =  1, diff =  -0.00
m =  2, diff =  +0.00
m =  3, diff =  +1.96
m =  4, diff =  +0.00
m =  5, diff = -13.69
m =  6, diff =  +1.96
m =  7, diff = -31.17
m =  8, diff =  +0.00
m =  9, diff =  +3.91
m = 10, diff = -13.69
m = 11, diff = -48.68
m = 12, diff =  +1.96
m = 13, diff = +40.53
m = 14, diff = -31.17
m = 15, diff = -11.73
m = 16, diff =  +0.00
```

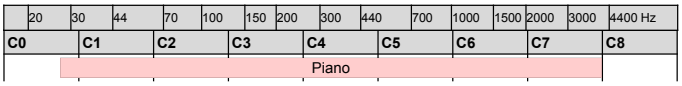
Exercise 1.10. Pythagorean tuning (named after the ancient Greek mathematician and philosopher Pythagoras) is a system of musical tuning in which the frequency ratios of all intervals are based on the ratio $3:2$ as found in the harmonic series. This ratio is also known as the **perfect fifth**. A **Pythagorean scale** is a scale constructed from only pure perfect fifths $(3:2)$ and octaves $(2:1)$. To obtain such a scale, start with the center frequency of the note C2, successively multiply the frequency value by a factor of $3/2$, and if necessary, divide it by two such that all frequency values lie between C2 and C3. Repeat this procedure to produce 13 frequency values (including the one for C2). As in Exercise 1.9, determine for each such frequency value the closest note of the equal-tempered scale (along with the difference in cents). The last of the produced frequency values is closest to the fundamental frequency of the note C3. The difference between the produced frequency and the center frequency of C3 is known as the **Pythagorean comma**, which indicates the degree of inconsistency when trying to define a twelve-tone scale using only perfect fifths.

Solution to Exercise 1.10. The following table yields the ratios of the Pythagorean tuning as well as the frequency ratios with regard of the twelve-tone equal-tempered scale of the closest notes.

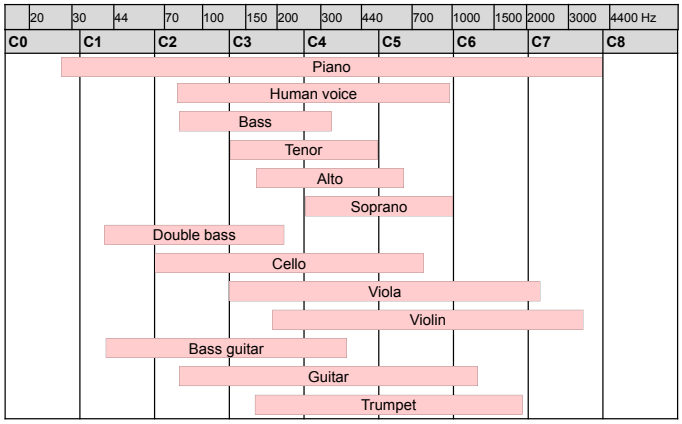
#	Pytagorean ratio			Equal-tempered scale			Difference (cents)
				Note	Frequency ratio		
0	1:1	1:1	1.0000	C2	1	1.0000	+0.00
1	3:2	3:2	1.5000	G2	$2^{7/12}$	1.4983	+1.96
2	3^2:2^3	9:8	1.1250	D2	$2^{2/12}$	1.1225	+3.91
3	3^3:2^4	27:16	1.6875	A2	$2^{9/12}$	1.6818	+5.87
4	3^4:2^6	81:64	1.2656	E2	$2^{4/12}$	1.2599	+7.82
5	3^5:2^7	243:128	1.8984	B3	$2^{11/12}$	1.8877	+9.78
6	3^6:2^9	729:512	1.4238	F♯2	$2^{6/12}$	1.4142	+11.73
7	3^7:2^{11}	2187:2048	1.0679	C♯2	$2^{1/12}$	1.0595	+13.69
8	3^8:2^{12}	6561:4096	1.6018	G♯2	$2^{8/12}$	1.5874	+15.64
9	3^9:2^{14}	19683:16384	1.2014	D♯2	$2^{3/12}$	1.1892	+17.60
10	3^{10}:2^{15}	59049:32768	1.8020	A♯2	$2^{10/12}$	1.7818	+19.55
11	3^{11}:2^{17}	177147:131072	1.3515	F2 (E♯2)	$2^{5/12}$	1.3348	+21.51
12	3^{12}:2^{19}	531441:524288	1.0136	C2 (B♯2)	1	1.0000	+23.46

The Pythagorean comma is the frequency ratio $3^{12}/2^{19} = 531441/524288 \approx 1.0136$, which corresponds to approximately 23.46 cents.

Exercise 1.11. Investigate the typical frequency range as well as pitch range of musical instruments (including the human voice) and graphically display this information as indicated by the following figure. For example, consider the ranges of standard instruments as used in Western orchestras including the piano, human voice (bass, tenor, alto, soprano), double bass, cello, viola, violin, bass guitar, guitar, trumpet. Similarly, consider instruments you are familiar with.



Solution to Exercise 1.11.



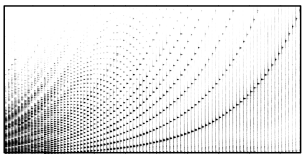
Exercise 1.12. Suppose that the intensity of a sound has been increased by 17 dB as defined in (1.6). Determine the factor by which the sound intensity has been increased.

Solution to Exercise 1.12. Let I_{ref} be the reference sound intensity and I be the sound intensity increased by 17 dB. By (1.6), this means that

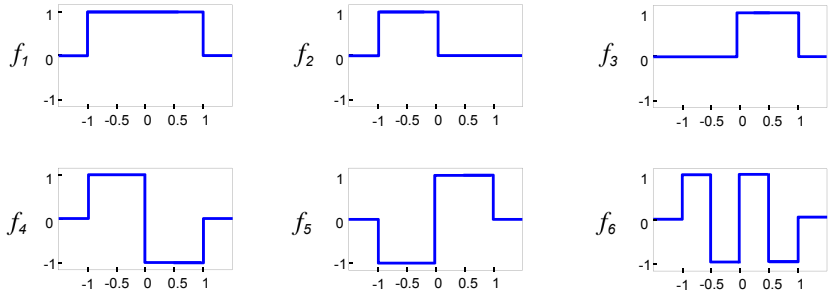
$$17 = 10 \cdot \log_{10}\left(\frac{I}{I_{\text{ref}}}\right).$$

Therefore the I differs from I_{ref} by a factor of $10^{17/10} \approx 50.119$.

Chapter 2
Fourier Analysis of Signals



Exercise 2.1. Let $\langle f|g\rangle := \int_{t\in\mathbb{R}} f(t)\cdot g(t)dt$ be the similarity measure for two functions $f:\mathbb{R}\to\mathbb{R}$ and $g:\mathbb{R}\to\mathbb{R}$ as defined in (2.3). Consider the following six functions $f_n:\mathbb{R}\to\mathbb{R}$ for $n\in[1:6]$, which are defined to be zero outside the shown interval:

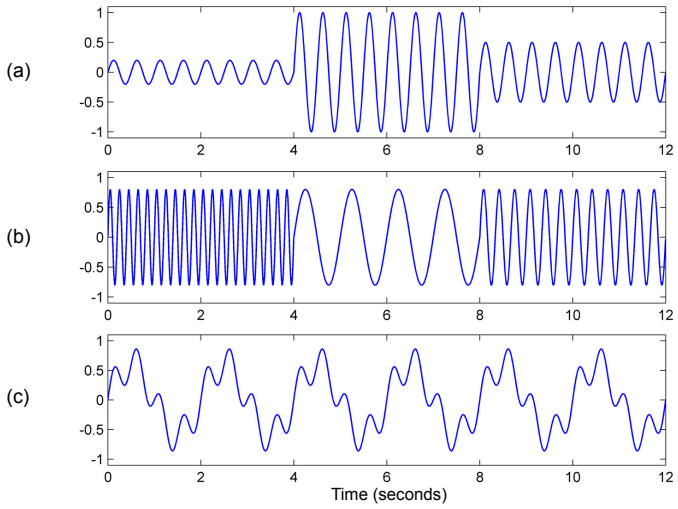


Determine the similarity values $\langle f_n|f_m\rangle$ for all pairs $(n,m)\in[1:6]\times[1:6]$.

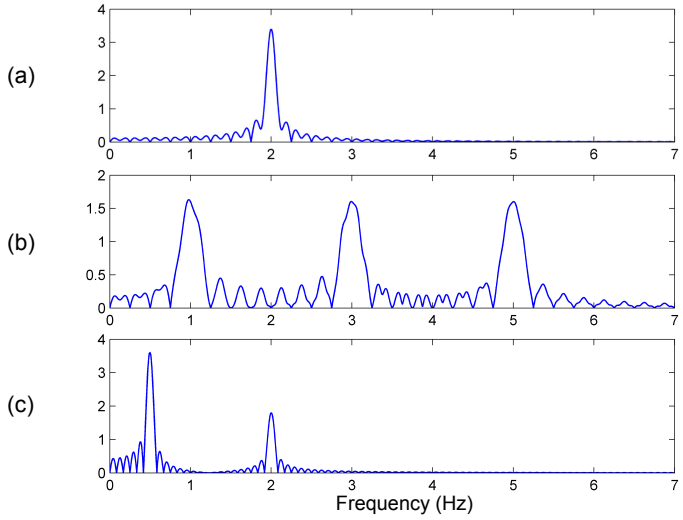
Solution to Exercise 2.1.

$\langle f_n\|f_m\rangle$	f_1	f_2	f_3	f_4	f_5	f_6
f_1	2	1	1	0	0	0
f_2	1	1	0	1	-1	0
f_3	1	0	1	-1	1	0
f_4	0	1	-1	2	-2	0
f_5	0	-1	1	-2	2	0
f_6	0	0	0	0	0	2

Exercise 2.2. Sketch the magnitude Fourier transform of the following signals assuming that the signals are zero outside the shown intervals (see Figure 2.6 for similar examples):



Solution to Exercise 2.2.



Exercise 2.3. Based on (2.27) and (2.28), compute the time resolution (in ms) and frequency resolution (in Hz) of a discrete STFT based on the following parameter settings:

(a) $F_s = 22050, N = 1024, H = 512$
(b) $F_s = 48000, N = 1024, H = 256$
(c) $F_s = 4000, N = 4096, H = 1024$

What are the respective Nyquist frequencies?

Solution to Exercise 2.3. The time resolution (in ms) is given by $1000 \cdot H/F_{\mathrm{s}}$, the frequency resolution (in Hz) by F_{s}/N, and the Nyquist frequency by $F_{\mathrm{s}}/2$. From this one obtains:

(a) Time resolution: 23.22 ms. Frequency resolution: 21.53 Hz. Nyquist frequency: 11025 Hz.

(b) Time resolution: 5.33 ms. Frequency resolution: 46.88 Hz. Nyquist frequency: 24000 Hz.

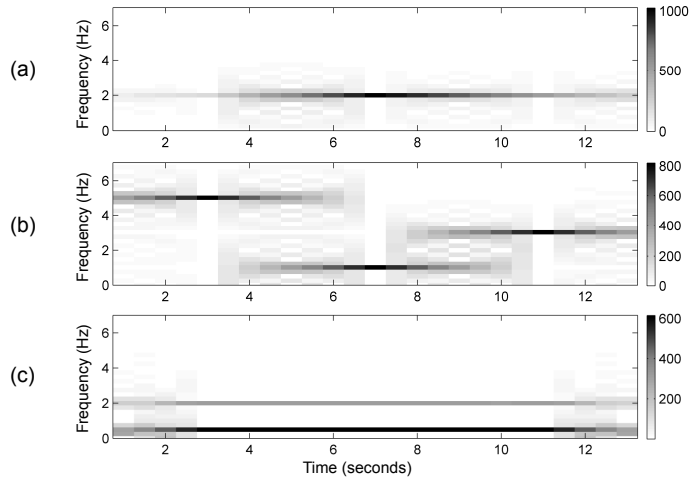
(c) Time resolution: 256.00 ms. Frequency resolution: 0.98 Hz. Nyquist frequency: 2000 Hz.

Exercise 2.4. Let $F_{\mathrm{s}} = 44100$, $N = 2048$, and $H = 1024$ be the parameter settings of a discrete STFT $\mathcal{X}$ as defined in (2.26). What is the physical meaning of the Fourier coefficients $\mathcal{X}(1000,1000)$, $\mathcal{X}(17,0)$, and $\mathcal{X}(56,1024)$, respectively? Why is the coefficient $\mathcal{X}(56,1024)$ problematic?

Solution to Exercise 2.4. According to (2.27) and (2.28), the coefficient $\mathcal{X}(1000,1000)$ corresponds to the physical time $T_{\mathrm{coef}}(m) = 23.22$ sec and the physical frequency $F_{\mathrm{coef}}(1000) = 21533$ Hz. Similarly, one obtains $T_{\mathrm{coef}}(17) = 0.39$ sec and $F_{\mathrm{coef}}(0) = 0$ Hz for $\mathcal{X}(17,0)$. Furthermore, one obtains $T_{\mathrm{coef}}(56) = 1.30$ sec and $F_{\mathrm{coef}}(1024) = 22050$ Hz for $\mathcal{X}(56,1024)$. The frequency expressed by the coefficient $\mathcal{X}(56,1024)$ corresponds to the Nyquist frequency. In general, this coefficient yields a poor approximation of the actual frequency of the underlying analog signal.

Exercise 2.5. Sketch the magnitude Fourier transform (as in Figure 2.9) for each of the three signals shown in Exercise 2.2. Assume a window length that corresponds to a physical duration of about one second.

Solution to Exercise 2.5.



Exercise 2.6. The naive approach for computing a DFT requires about N^2 operations, while the FFT requires about $N \log_2 N$ operations. Compute the factor for

the savings when using the FFT for various N. In particular, consider $N = 2^n$ for $n = 5, 10, 15, 20, 25, 30$.

Solution to Exercise 2.6. Let $N = 2^n$. The factor for the savings is $N^2/(N\log_2 N) = N/n$. The following table yields the factors (rounded to integers) for various $N = 2^n$:

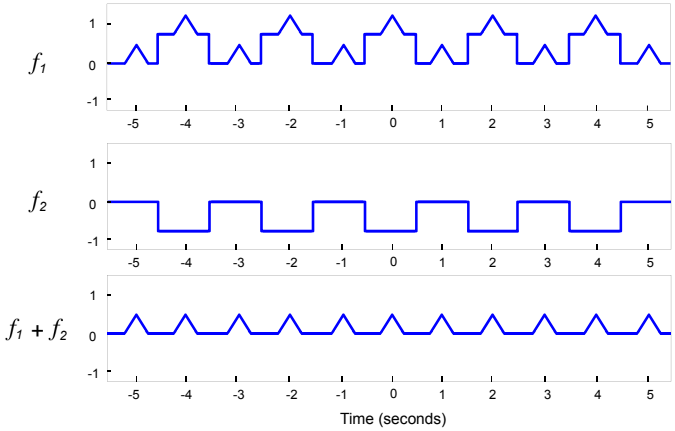
n	5	10	15	20	25	30
N/n	6	102	2185	52429	1342177	35791394

Exercise 2.7. Let f_1 and f_2 be two periodic analog signals with integer periods $\lambda_1 \in \mathbb{N}$ and $\lambda_2 \in \mathbb{N}$, respectively. Show that $g = f_1 + f_2$ is periodic with periods that are integer multiples of λ_1 as well as λ_2. In general, g may have additional periods not necessarily being integer multiples of λ_1 and λ_2. As an example, specify two signals f_1 and f_2 with prime period $\lambda_1 = \lambda_2 = 2$ such that $g = f_1 + f_2$ is periodic with prime period $\lambda = 1$.

Solution to Exercise 2.7. First note that a periodic function f with period λ is also periodic with period $n\lambda$ for any integer $n \in \mathbb{N}$. Now, let $\lambda \in \mathbb{Z}$ be a number that is an integer multiple of λ_1 as well as of λ_2. Then there exist integer numbers $n_1, n_2 \in \mathbb{Z}$ such that $\lambda = n_1\lambda_1 = n_2\lambda_2$ and

$$\begin{aligned} g(t+\lambda) &= f_1(t+\lambda) + f_2(t+\lambda) \\ &= f_1(t+n_1\lambda_1) + f_2(t+n_2\lambda_2) \\ &= f_1(t) + f_2(t) = g(t). \end{aligned}$$

This shows that g is periodic for each such period λ. The following example shows that two functions f_1 and f_2 with prime period 2 may sum up to some function $g = f_1 + f_2$ with prime period 1:



Exercise 2.8. In this exercise, we show that there are periodic functions that do not have a prime period (i.e., that do not have a least positive constant being a period). The easiest example of such a function is a constant function. Show that the function $f : \mathbb{R} \to \mathbb{R}$ defined by

$$f(t) := \begin{cases} 1, & \text{for } t \in \mathbb{Q}, \\ 0, & \text{for } t \in \mathbb{R} \setminus \mathbb{Q} \end{cases}$$

is also periodic without having a prime period.
[**Hint:** In this exercise, we assume that the reader is familiar with the properties of rational numbers ($\mathbb{Q}$) and irrational numbers ($\mathbb{R} \setminus \mathbb{Q}$).]

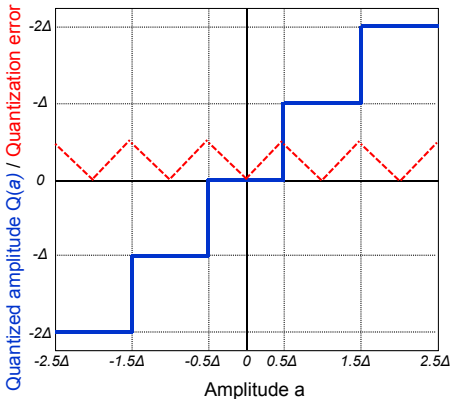
Solution to Exercise 2.8. Adding a rational number to another rational number yields a rational number. Furthermore, adding a rational number to an irrational number yields an irrational number. Therefore, $f(t+q) = f(t)$ for any rational number $q \in Q$. This shows that f is periodic with regard to any rational number $q \in Q$. Since there are arbitrarily small rational numbers, f has no prime period.

Exercise 2.9. Sketch the graph of the quantization function $Q : \mathbb{R} \to \mathbb{R}$ defined by

$$Q(a) := \operatorname{sgn}(a) \cdot \Delta \cdot \left\lfloor \frac{|a|}{\Delta} + \frac{1}{2} \right\rfloor$$

for $a \in \mathbb{R}$ and some fixed quantization step size $\Delta > 0$ (see (2.33)). Furthermore, sketch the graph of the absolute quantization error.

Solution to Exercise 2.9.



Exercise 2.10. In mathematics, the term "operator" is used to denote a mapping from one vector space to another. Let V and W be two vector spaces over $\mathbb{R}$. An operator $M : V \to W$ is called **linear** if $M[a_1 v_1 + a_2 v_2] = a_1 M[v_1] + a_2 M[v_2]$ for any $v_1, v_2 \in V$ and $a_1, a_2 \in \mathbb{R}$. Show that $V := \{f \mid f\colon \mathbb{R} \to \mathbb{R}\}$ and $W := \{x \mid x\colon \mathbb{Z} \to \mathbb{R}\}$ are vector spaces. Fixing a sampling period $T > 0$, consider the operator M that maps a CT-signal $f \in V$ to the DT-signal $M[f] := x \in W$ obtained by T-sampling as defined in (2.32). Show that this defines a linear operator.

Solution to Exercise 2.10. For two CT-signals $f_1, f_2 \in V$ and real numbers $a_1, a_2 \in \mathbb{R}$, one obtains a CT-signal $a_1 f_1 + a_2 f_2 \in V$ by (2.30) and (2.31). This shows that V is a vector space. Similar definitions show that W is a vector space. Next, let $x_1 := M[f_1]$ and $x_2 := M[f_2]$ be the DT-signals obtained by T-sampling. From

$$\begin{aligned} M[a_1 f_1 + a_2 f_2](n) &= (a_1 f_1 + a_2 f_2)(n \cdot T) \\ &= a_1 f_1(n \cdot T) + a_2 f_2(n \cdot T) \\ &= a_1 M[f_1](n) + a_2 M[f_2](n) \\ &= (a_1 M[f_1] + a_2 M[f_2])(n) \end{aligned}$$

for all $n \in \mathbb{Z}$, it follows that $M(a_1 f_1 + a_2 f_2) = a_1 M(f_1) + a_2 M(f_2)$. This proves that T-sampling is a linear operator.

Exercise 2.11. Show that the quantization operator $Q : \mathbb{R} \to \mathbb{R}$ as defined in Exercise 2.9 and (2.33) is *not* a linear operator.

Solution to Exercise 2.11. For example, one has $4 \cdot Q(0.25) = 0 \neq 1 = Q(1) = Q(4 \cdot 0.25)$. This shows that Q is not linear.

Exercise 2.12. In this exercise we discuss various computation rules for complex numbers and their conjugates. The complex multiplication is defined by $c_1 \cdot c_2 = a_1 a_2 - b_1 b_2 + i(a_1 b_2 + a_2 b_1)$ for two complex numbers $c_1 = a_1 + ib_1, c_2 = a_2 + ib_2 \in \mathbb{C}$ (see (2.34)). Furthermore, complex conjugation is defined by $\overline{c} = a - ib$ for a complex number $c = a + ib \in \mathbb{C}$ (see (2.35)). Finally, the absolute value of a complex number c is defined by $|c| = \sqrt{a^2 + b^2}$. Prove the following identities:

(a) $\mathrm{Re}(c) = (c + \overline{c})/2$
(b) $\mathrm{Im}(c) = (c - \overline{c})/(2i)$
(c) $\overline{c_1 + c_2} = \overline{c_1} + \overline{c_2}$
(d) $\overline{c_1 \cdot c_2} = \overline{c_1} \cdot \overline{c_2}$
(e) $c\overline{c} = a^2 + b^2 = |c|^2$
(f) $1/c = \overline{c}/(c\overline{c}) = \overline{c}/(a^2 + b^2) = \overline{c}/(|c|^2)$

Solution to Exercise 2.12.

(a) Follows from $c + \overline{c} = a + ib + a - ib = 2a = 2\mathrm{Re}(c)$.
(b) Follows from $c - \overline{c} = a + ib - a + ib = 2ib = 2i\mathrm{Im}(c)$.
(c) $\overline{c_1 + c_2} = (a_1 + a_2) - i(b_1 + b_2) = (a_1 - ib_1) + (a_2 - ib_2) = \overline{c_1} + \overline{c_2}$
(d) $\overline{c_1 \cdot c_2} = a_1 a_2 - b_1 b_2 - i(a_1 b_2 + a_2 b_1) = (a_1 - ib_1)(a_2 - ib_2) = \overline{c_1} \cdot \overline{c_2}$
(e) $c\overline{c} = (a + ib)(a - ib) = a^2 + b^2 + i(-ab + ba) = a^2 + b^2 = |c|^2$
(f) Follows from $1 = c\overline{c}/(c\overline{c}) = c \cdot (\overline{c}/(c\overline{c}))$ and (e).

Exercise 2.13. We have seen in Section 2.2.3.2 that the set $\mathbb{C}^{\mathbb{Z}} = \{x | x\colon \mathbb{Z} \to \mathbb{C}\}$ of complex-valued DT-signals defines a vector space. Show that the subset $\ell^2(\mathbb{Z}) \subset \mathbb{C}^{\mathbb{Z}}$ of DT-signals of finite energy is a linear subspace. To this end, you need to show that $x + y \in \ell^2(\mathbb{Z})$ and $ax \in \ell^2(\mathbb{Z})$ for any $x, y \in \ell^2(\mathbb{Z})$ and $a \in \mathbb{C}$.

Solution to Exercise 2.13. Let $x, y \in \ell^2(\mathbb{Z})$ and $a \in \mathbb{C}$. By definition (2.42), one has $\mathrm{E}(x) < \infty$ and $\mathrm{E}(y) < \infty$. This implies $\mathrm{E}(ax) = |a|^2 \mathrm{E}(x) < \infty$, i.e., $ax \in \ell^2(\mathbb{Z})$. Furthermore,

$$\begin{aligned}\mathrm{E}(x+y) &= \sum_{n\in\mathbb{Z}} |x(n)+y(n)|^2 \;\le\; \sum_{n\in\mathbb{Z}} (|x(n)|+|y(n)|)^2 \\ &\le \sum_{n\in\mathbb{Z}} (2\max\{|x(n)|,|y(n)|\})^2 \;\le\; 4\sum_{n\in\mathbb{Z}} (|x(n)|^2+|y(n)|^2) \\ &= 4(\mathrm{E}(x)+\mathrm{E}(y)) < \infty.\end{aligned}$$

This shows that $x+y \in \ell^2(\mathbb{Z})$.

Exercise 2.14. In Section 2.3.1, we defined the set $\{\mathbf{1}, \mathbf{sin}_k, \mathbf{cos}_k \mid k \in \mathbb{N}\} \subset L^2_{\mathbb{R}}([0,1))$. Prove that this set is an orthonormal set in $L^2_{\mathbb{R}}([0,1))$, i.e., that it satisfies (2.50) and (2.51).
[**Hint:** Use the following trigonometric identities:

(a) $\cos(\alpha)^2 + \sin(\alpha)^2 = 1$
(b) $\cos(\alpha)\cos(\beta) = (\cos(\alpha+\beta)+\cos(\alpha-\beta))/2$
(c) $\sin(\alpha)\sin(\beta) = (\cos(\alpha-\beta)-\cos(\alpha+\beta))/2$
(d) $\sin(\alpha)\cos(\beta) = (\sin(\alpha+\beta)+\sin(\alpha-\beta))/2$

To show (2.51), use (a) and the fact that $\mathbf{cos}_k^2$ and $\mathbf{sin}_k^2$ have the same area over a full period. The proof of (2.50) is a bit cumbersome, but not difficult when using (b), (c), and (d).]

Solution to Exercise 2.14. First, we prove (2.51). Obviously, one has $\|\mathbf{1}\|^2 = 1$. Furthermore, from identity (a), one obtains

$$2 = 2(\cos(2\pi kt)^2 + \sin(2\pi kt)^2) = \cos_k(t)^2 + \sin_k(t)^2$$

for all $t \in [0,1)$. Therefore,

$$2 = \int_{t\in[0,1)} \cos_k(t)^2 + \sin_k(t)^2 dt = \langle \cos_k | \cos_k \rangle + \langle \sin_k | \sin_k \rangle.$$

Both functions $\cos_k^2$ and $\sin_k^2$ are 1-periodic and shifted versions from each other. Therefore, integration of both functions over a full period yields the same value. As a result, one obtains $\langle \cos_k | \cos_k \rangle = \langle \sin_k | \sin_k \rangle = 1$.

Next, we prove (2.50).

$$\begin{aligned}\langle \mathbf{1} | \mathbf{cos}_k \rangle &= \int_{t\in[0,1)} \sqrt{2}\cos(2\pi kt) dt = \Big[\sqrt{2}\sin(2\pi kt)/(2\pi k)\Big]_0^1 = 0 \\ \langle \mathbf{1} | \mathbf{sin}_k \rangle &= \int_{t\in[0,1)} \sqrt{2}\sin(2\pi kt) dt = \Big[-\sqrt{2}\cos(2\pi kt)/(2\pi k)\Big]_0^1 = 0\end{aligned}$$

Using (b), one obtains for $k \neq \ell$:

$$\begin{aligned}\langle \mathbf{cos}_k | \mathbf{cos}_\ell \rangle &= \int_{t\in[0,1)} \sqrt{2}\cos(2\pi kt)\sqrt{2}\cos(2\pi \ell t)dt \\ &= 2\int_{t\in[0,1)} \frac{\cos(2\pi(k+\ell)t)+\cos(2\pi(k-\ell)t)}{2}dt \\ &= \left[\frac{\sin(2\pi(k+\ell)t)}{2\pi(k+\ell)} + \frac{\sin(2\pi(k-\ell)t)}{2\pi(k-\ell)}\right]_0^1 = 0\end{aligned}$$

Similarly, using (c), one shows $\langle \mathbf{sin}_k | \mathbf{sin}_\ell \rangle = 0$ for $k \neq \ell$. For $k \not\ell$, one obtains

$$\begin{aligned}\langle \mathbf{cos}_k | \mathbf{sin}_\ell \rangle &= \int_{t\in[0,1)} \sqrt{2}\cos(2\pi kt)\sqrt{2}\sin(2\pi \ell t)dt \\ &= 2\int_{t\in[0,1)} \frac{\sin(2\pi(k+\ell)t)+\sin(2\pi(k-\ell)t)}{2}dt \\ &= \left[\frac{\cos(2\pi(k+\ell)t)}{2\pi(k+\ell)} + \frac{\cos(2\pi(k-\ell)t)}{2\pi(k-\ell)}\right]_0^1 = 0.\end{aligned}$$

Finally, for $k = \ell$, one obtains

$$\begin{aligned}\langle \mathbf{cos}_k | \mathbf{sin}_k \rangle &= \int_{t\in[0,1)} \sqrt{2}\cos(2\pi kt)\sqrt{2}\sin(2\pi kt)dt \\ &= 2\int_{t\in[0,1)} \frac{\sin(2\pi(2k)t)}{2}dt = \left[\frac{-\cos(2\pi(2k)t)}{2\pi(k+\ell)}\right]_0^1 = 0.\end{aligned}$$

This concludes the proof.

Exercise 2.15. Let $\exp(i\gamma) := \cos(\gamma) + i\sin(\gamma)$, $\gamma \in \mathbb{R}$, be the complex exponential function as defined in (2.67). Prove the following properties (see (2.68) to (2.71)):

(a) $\exp(i\gamma) = \exp(i(\gamma+2\pi))$
(b) $|\exp(i\gamma)| = 1$
(c) $\overline{\exp(i\gamma)} = \exp(-i\gamma)$
(d) $\exp(i(\gamma_1+\gamma_2)) = \exp(i\gamma_1)\exp(i\gamma_2)$
(e) $\dfrac{d\exp(i\gamma)}{d\gamma} = i\exp(i\gamma)$

[**Hint:** To prove (d), you need the trigonometric identities $\cos(\alpha+\beta) = \cos(\alpha)\cos(\beta) - \sin(\alpha)\sin(\beta)$ and $\sin(\alpha+\beta) = \cos(\alpha)\sin(\beta) + \sin(\alpha)\cos(\beta)$. In (e), note that the real (imaginary) part of a derivative of a complex-valued function is obtained by computing the derivative of the real (imaginary) part of the function.]

Solution to Exercise 2.15. Property (a) follows from

$$\exp(i\gamma) = \cos(\gamma) + i\sin(\gamma) = \cos(\gamma+2\pi) + i\sin(\gamma+2\pi) = \exp(i(\gamma+2\pi)).$$

Using $\cos(\alpha)^2 + \sin(\alpha)^2 = 1$, property (b) follows from

$$|\exp(i\gamma)| = \sqrt{\cos(\gamma)^2 + \sin(\gamma)^2} = 1.$$

Using $\cos(\alpha) = \cos(-\alpha)$ and $\sin(\alpha) = -\sin(-\alpha)$, property (c) follows from

$$\overline{\exp(i\gamma)} = \cos(\gamma) - i\sin(\gamma) = \cos(-\gamma) + i\sin(-\gamma) = \exp(-i\gamma).$$

Using the two trigonemetric identities specified in the hint, property (d) follows from

$$\begin{aligned}\exp(i(\gamma_1+\gamma_2)) &= \exp(i\gamma_1)\exp(i\gamma_2)\\ &= \cos(\gamma_1+\gamma_2) + i\sin(\gamma_1+\gamma_2)\\ &= \cos(\gamma_1)\cos(\gamma_2) - \sin(\gamma_1)\sin(\gamma_2) + i(\cos(\gamma_1)\sin(\gamma_2) + \sin(\gamma_1)\cos(\gamma_2))\\ &= (\cos(\gamma_1) + i\sin(\gamma_1))\cdot(\cos(\gamma_2) + i\sin(\gamma_2))\\ &= \exp(i\gamma_1)\exp(i\gamma_2).\end{aligned}$$

The property (e) follows from

$$\begin{aligned}\frac{d\exp(i\gamma)}{d\gamma} &= \frac{d\cos(\gamma)}{d\gamma} + i\frac{d\sin(\gamma)}{d\gamma} = -\sin(\gamma) + i\cos(\gamma)\\ &= i(\cos(\gamma) + i\sin(\gamma)) = i\exp(i\gamma).\end{aligned}$$

Exercise 2.16. In (2.77), we defined for each $k \in \mathbb{Z}$ the complex-valued exponential function $\mathbf{exp}_k : [0,1) \to \mathbb{C}$ by $\mathbf{exp}_k(t) := \cos(2\pi kt) + i\sin(2\pi kt)$, $t \in \mathbb{R}$. As in Exercise 2.14, show that the set $\{\mathbf{exp}_k \mid k \in \mathbb{Z}\} \subset L^2([0,1))$ is an orthonormal set, i.e., $\|\mathbf{exp}_k\|^2 = 1$ for $k \in \mathbb{Z}$ (see (2.51)) and $\langle \mathbf{exp}_k | \mathbf{exp}_\ell \rangle = 0$ for $k \neq \ell$, $k, \ell \in \mathbb{Z}$ (see (2.50)).
[**Hint:** Use the properties of the exponential function introduced in Exercise 2.15. Furthermore, note that the real (imaginary) part of an integral of a complex-valued function is obtained by integrating the real (imaginary) part of the function.]

Solution to Exercise 2.16. By (2.47) and the property $|\exp(i\gamma)| = 1$, we obtain

$$\|\mathbf{exp}_k\|^2 = \mathrm{E}_{[0,1)}(\mathbf{exp}_k) = \int_{t\in[0,1)} |\exp(2\pi kt)|^2 dt = 1.$$

Next, let $k \neq \ell$. Then, from (2.49) and the properties in Exercise 2.15, we obtain

$$\begin{aligned}\langle \mathbf{exp}_k | \mathbf{exp}_\ell \rangle &= \int_{t\in[0,1)} \exp(2\pi ikt)\overline{\exp(2\pi i\ell t)} dt = \int_{t\in[0,1)} \exp(2\pi i(k-\ell)t) dt\\ &= \left[\frac{\exp(2\pi i(k-\ell)t)}{2\pi i(k-\ell)}\right]_0^1 = 0.\end{aligned}$$

Exercise 2.17. Let atan2 be the function as defined in (2.76). For a complex number $c = a + ib \in \mathbb{C}$, we set $\mathrm{atan2}(c) := \mathrm{atan2}(b,a)$. Show that $\mathrm{atan2}(\lambda \cdot c) = \mathrm{atan2}(c)$ for any positive constant $\lambda \in \mathbb{R}_{>0}$. Furthermore, show that $\mathrm{atan2}(\bar{c}) = -\mathrm{atan2}(c)$.
[**Hint:** Use the fact that the arctan function is an odd function, i.e., $\arctan(-v) = -\arctan(v)$ for $v \in \mathbb{R}$.]

Solution to Exercise 2.17. When considering $\lambda \cdot c$ instead of c, the six cases concerning the relation between a and b in the definition of (2.76) do not change. Furthermore, $\lambda b/\lambda a = b/a$. From this, one obtains $\operatorname{atan2}(\lambda \cdot c) = \operatorname{atan2}(c)$.

Furthermore, when considering $\overline{c}$ instead of c, one has $-b$ instead of b. Therefore, the cases two and three in the definition of (2.76) are interchanged. Suppose $b \geq 0$, then

$$\begin{aligned}\operatorname{atan2}(-b/a) &= \arctan(-b/a) - \pi = -\arctan(b/a) - \pi \\ &= -(\arctan(b/a) + \pi) = -\operatorname{atan2}(b/a).\end{aligned}$$

Suppose $b < 0$, then

$$\begin{aligned}\operatorname{atan2}(-b/a) &= \arctan(-b/a) + \pi = -\arctan(b/a) - \pi \\ &= -(\arctan(b/a) - \pi) = -\operatorname{atan2}(b/a).\end{aligned}$$

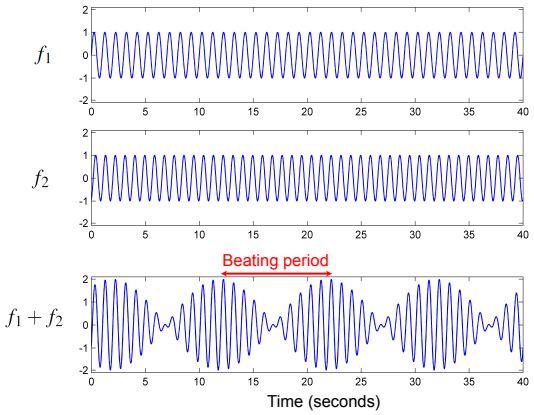
Furthermore, the cases four and five in the definition of (2.76) are interchanged. Again, one obtains $\operatorname{atan2}(-b/a) = -\operatorname{atan2}(b/a)$. Altogehter, we have shown that $\operatorname{atan2}(\overline{c}) = -\operatorname{atan2}(c)$.

Exercise 2.18. In this exercise, we consider the geometric series for compex numbers, which is needed in (2.112). Prove that $\sum_{n=0}^{N-1} a^n = (1-a^N)/(1-a)$ for any complex number $a \neq 1$.
[**Hint:** For the proof, use mathematical induction on N.]

Solution to Exercise 2.18. For $N = 1$, one obtains $\sum_{n=0}^{0} a^n = 1 = a^0 = (1-a)/(1-a)$ and the assertion is true. Now, let $N > 1$ and assume that the assertion is true for $N-1$. Then we obtain

$$\begin{aligned}\sum_{n=0}^{N-1} a^n &= a^{N-1} + \sum_{n=0}^{N-2} a^n = a^{N-1} + \frac{1-a^{N-1}}{1-a} \\ &= \frac{a^{N-1} - a^N + 1 - a^{N-1}}{1-a} = \frac{1-a^N}{1-a}.\end{aligned}$$

Exercise 2.19. We have seen that two sinusoids of similar frequency may add up (constructive interference) or cancel out (destructive interference); see Figure 2.19. Let $f_1(t) = \sin(2\pi\omega_1 t)$ and $f_2(t) = \sin(2\pi\omega_2 t)$ be two such sinusoids with distinct but nearby frequencies $\omega_1 \approx \omega_2$. In the following figure, for example, $\omega_1 = 1$ and $\omega_2 = 1.1$ is used.



The figure also shows that the superposition $f_1 + f_2$ of these two sinusoids results in a function that looks like a single sine wave with a slowly varying amplitude, a phenomenon also known as *beating*. Determine the rate (reciprocal of the period) of the beating in dependency on ω_1 and ω_2. Compare this result with the plot of $f_1 + f_2$ in the figure.

[**Hint:** Use the trigonometric identity $\sin(\alpha) + \sin(\beta) = 2\cos\left(\frac{\alpha-\beta}{2}\right)\sin\left(\frac{\alpha+\beta}{2}\right)$ for $\alpha, \beta \in \mathbb{R}$.]

Solution to Exercise 2.19. Setting $\alpha = 2\pi\omega_1 t$ and $\beta = 2\pi\omega_2 t$, one obtains:

$$\sin(2\pi\omega_1 t) + \sin(2\pi\omega_2 t) = 2\cos\left(2\pi\frac{\omega_1 - \omega_2}{2}t\right)\sin\left(2\pi\frac{\omega_1 + \omega_2}{2}t\right)$$

This shows that if the difference $\omega_1 - \omega_2$ is small, the cosine term has a low frequency compared with the sine term. As a result the signal $f_1 + f_2$ can be seen as a sine wave of frequency $(\omega_1 + \omega_2)/2$ with a slowly varying amplitude envelope of frequency $|\omega_1 - \omega_2|$. Note that this rate is twice the frequency $(\omega_1 - \omega_2)/2$ of the cosine term. In the example with $\omega_1 = 1$ and $\omega_2 = 1.1$ shown in the figure, the beating rate is 0.1 Hz and the beating period is 10 sec.

Exercise 2.20. Let $f \in L^2(\mathbb{R})$ be a signal of unit energy $\|f\|^2 = 1$. Show that the scaled signal g defined by $g(t) := s^{1/2} f(s \cdot t)$ also has unit energy for a positive real scaling factor $s > 0$. Furthermore show that $\hat{g}(\omega) = s^{-1/2}\hat{f}(\omega/s)$ for $\omega \in \mathbb{R}$. Discuss this result. Describe how one can obtain a Dirac sequence by changing the parameter s (see Section 2.3.3.2).

Solution to Exercise 2.20. The assertion $\|g\|^2 = 1$ follows from

$$\|g\|^2 = \int_{t\in\mathbb{R}} |g(t)|^2 dt = \int_{t\in\mathbb{R}} s|f(st)|^2 dt = \int_{t\in\mathbb{R}} |f(t)|^2 dt = \|f\|^2.$$

For the Fourier transform $\hat{g}$ holds:

$$\hat{g}(\omega) = \int_{t\in\mathbb{R}} g(t)\exp(-2\pi i\omega t)dt \;=\; \int_{t\in\mathbb{R}} s^{1/2} f(st)\exp(-2\pi i\omega t)dt$$
$$= \int_{t\in\mathbb{R}} s^{-1/2} f(t)\exp(-2\pi i\omega t/s)dt \;=\; s^{-1/2}\hat{f}(\omega/s).$$

For increasing s, the function g becomes narrower and the function $\hat{g}$ wider. For example, if f is the Gaussian as defined in (2.94), one obtains a Dirac sequence when using an increasing sequence of scaling factors approaching infinity.

Exercise 2.21. Show that the Fourier transform of the rectangular function in (2.95) is the sinc function in (2.96). Also prove that the sinc function is continuous at $t = 0$. [**Hint:** Use the fact that the derivative of $t \mapsto \exp(-2\pi i\omega t)$ is given by $t \mapsto -2\pi i\omega\exp(-2\pi i\omega t)$; see Exercise 2.15. From this, one can derive the indefinite integral of the exponential function. To prove the continuity at $t = 0$, look at the first terms of the Taylor series of the sine function.]

Solution to Exercise 2.21. For $\omega \neq 0$ one obtains

$$\hat{f}(\omega) = \int_{t\in\mathbb{R}} f(t)\exp(-2\pi i\omega t)dt \;=\; \int_{-1/2}^{1/2} \exp(-2\pi i\omega t)dt$$
$$= \left[\frac{1}{-2\pi i\omega}\exp(-2\pi i\omega t)\right]_{-1/2}^{1/2}$$
$$= \frac{1}{-2\pi i\omega}\left(\exp(-\pi i\omega) - \exp(\pi i\omega)\right)$$
$$= \frac{1}{2\pi i\omega}\left(\exp(\pi i\omega) - \overline{\exp(\pi i\omega)}\right) \quad = \quad \frac{\sin(\pi\omega)}{\pi\omega}.$$

For $\omega = 0$ we get

$$\hat{f}(0) = \int_{-\infty}^{\infty} f(t)dt = 1.$$

This shows that the derivative of the rectangular function is the sinc function. Furthermore, from $\sin(t) = t - \frac{t^3}{3!} + \frac{t^5}{5!} \mp \ldots$ follows that $\lim_{t\to 0}\frac{\sin(t)}{t} = 1$, which proves the continuity of the sinc function.

Exercise 2.22. For a signal $f \in L^2(\mathbb{R})$, consider the translation f_{t_0} defined by $f_{t_0}(t) := f(t - t_0)$ for $t \in \mathbb{R}$ (see (2.97)) and the modulation f^{ω_0} defined by $f^{\omega_0}(t) := \exp(2\pi i\omega_0 t)f(t)$ for $t \in \mathbb{R}$ (see (2.98)). Show that $\|f\| = \|f_{t_0}\| = \|f^{\omega_0}\|$. Furthermore, prove the properties (2.99) and (2.100):

$$\widehat{f_{t_0}}(\omega) = \exp(-2\pi i\omega t_0)\hat{f}(\omega) \qquad \text{and} \qquad \widehat{f^{\omega_0}}(\omega) = \hat{f}(\omega + \omega_0)$$

for $\omega \in \mathbb{R}$.

Solution to Exercise 2.22. The identity $\|f\| = \|f_{t_0}\|$ follows from a simple substitution of $t - t_0$ by t in the integration. The identity $\|f\| = \|f^{\omega_0}\|$ follows from $|\exp(2\pi i\omega_0 t)| = 1$. Furthermore,

$$\begin{aligned}\widehat{f_{t_0}}(\omega) &= \int_{t\in\mathbb{R}} f(t-t_0)\exp(2\pi i\omega t)dt \\ &= \int_{t\in\mathbb{R}} f(t)\exp(2\pi i\omega(t+t_0))dt \\ &= \exp(2\pi i\omega t_0)\int_{t\in\mathbb{R}} f(t)\exp(2\pi i\omega t)dt \\ &= \exp(2\pi i\omega t_0)\hat{f}(\omega).\end{aligned}$$

Finally, one obtains

$$\begin{aligned}\widehat{f^{\omega_0}}(\omega) &= \int_{t\in\mathbb{R}} \exp(2\pi i\omega_0 t)f(t)\exp(2\pi i\omega t)dt \\ &= \int_{t\in\mathbb{R}} f(t)\exp(2\pi i(\omega+\omega_0)t)dt \\ &= \hat{f}(\omega+\omega_0).\end{aligned}$$

Exercise 2.23. Any complex number $c\in\mathbb{C}$ with $c^N = 1$ for a given $N\in\mathbb{N}$ is called an N^{th} **root of unity**. If in addition $c^k \neq 1$ for $1 < k < N$, the root c is called **primitive**. Show that $\rho_N := \exp(-2\pi i/N)$ defines a primitive N^{th} root of unity. Furthermore, describe *all* N^{th} roots of unity. Which of these roots are primitive? Determine for $N\in\{4,7,12\}$ all primitive N^{th} roots of unity.
[**Hint:** In this exercise, one needs to know that a (nonzero) polynomial of degree N has at most N different roots, where a **root** of a function is an input value that produces an output of zero.]

Solution to Exercise 2.23. For $\rho_N := \exp(-2\pi i/N)$, we consider the powers $\rho_N^k = \exp(-2\pi ik/N)$ for $k\in[0:N-1]$. Suppose $\rho_N^k = \rho_N^\ell$ for some $k,\ell\in[0:N-1]$ with $k\geq\ell$. Then $1 = \rho_N^{(k-\ell)} = \exp(-2\pi i(k-\ell)/N)$ implies that $((k-\ell)/N)\in\mathbb{Z}$. Since $(k-\ell) < N$, this is only possible for $k=\ell$. This shows that the numbers ρ_N^k, $k\in[0:N-1]$, are all distinct and that ρ_N is primitive.

Furthermore, $(\rho_N^k)^N = \exp(-2\pi ik/N)^N = \exp(-2\pi ik) = 1$. Since the polynomial X^N-1 has at most N distinct roots, there can be at most N roots of unity. Therefore, the numbers ρ_N^k, $k\in[0:N-1]$, cover all N^{th} roots of unity.

Let us now fix a $k\in[0:N-1]$. Suppose that $(\rho_N^k)^\ell = 1$ for some $\ell\in[1:N-1]$. This is equivalent to $\exp(-2\pi ik\ell/N) = 1$ or $k\ell/N\in\mathbb{Z}$. In other words, the product $k\ell$ is then divisible by N. Since $0 < \ell < N$, this is equivalent for k and N having a common divisor larger than 1.

From this, we obtain the following primitive roots of unity. For $N=4$, ρ_N^k for $k\in\{1,3\}$. For $N=7$, ρ_N^k for $k\in\{1,2,3,4,5,6\}$. For $N=12$, ρ_N^k for $k\in\{1,5,7,11\}$.

Exercise 2.24. Let $\mathbf{x} = (x(0),\ldots x(N-1))^\top$ be a real-valued vector consisting of samples $x(n)\in\mathbb{R}$ for $n\in[0:N-1]$. Show that

$$\mathbf{X} = \mathrm{DFT}_N\cdot\mathbf{x}$$

with $\mathbf{X} = (X(0),\ldots X(N-1))^\top$ fulfills the symmetry property $X(k) = \overline{X(N-k)}$ for all $k\in[1:N-1]$ and $X(0)\in\mathbb{R}$. This shows that the upper half of the frequency

coefficients are redundant if $\mathbf{x}$ is real-valued. Furthermore, show the converse. Given a spectral vector $\mathbf{X}$ with $X(0) \in \mathbb{R}$ and $X(k) = \overline{X(N-k)}$ for all $k \in [1:N-1]$, then

$$\mathbf{x} = \mathrm{DFT}_N^{-1} \cdot \mathbf{X}$$

is a real-valued vector (see (2.118)).
[**Hint:** Use the computation rules for complex numbers from Exercise 2.12.]

Solution to Exercise 2.24. Let $\mathbf{x}$ be real-valued, then $x(n) = \overline{x(n)}$ for all $n \in [0:N-1]$. From the computation rules for complex numbers (see Exercise 2.12) and (2.70) follows:

$$\begin{aligned}
\overline{X(N-k)} &= \overline{\sum_{n=0}^{N-1} x(n) \exp(-2\pi i(N-k)n/N)} \\
&= \sum_{n=0}^{N-1} \overline{x(n)} \cdot \overline{\exp(-2\pi i n)} \cdot \overline{\exp(2\pi i k n/N)} \\
&= \sum_{n=0}^{N-1} x(n) \exp(-2\pi i k n/N) \\
&= X(k)
\end{aligned}$$

for $k \in [1:N-1]$. Furthermore $X(0) = \sum_{n=0}^{N-1} x(n) \in \mathbb{R}$ in case all samples are real-valued. Now, let us suppose that we are given a spectral vector $\mathbf{X}$ with $X(0) \in \mathbb{R}$ and $X(k) = \overline{X(N-k)}$ for all $k \in [1:N-1]$. Then, using (2.118), we obtain

$$\begin{aligned}
\overline{x(n)} &= \frac{1}{N} \overline{\sum_{k=0}^{N-1} X(k) \exp(2\pi i n k/N)} \\
&= \frac{1}{N} \sum_{k=0}^{N-1} \overline{X(k)} \exp(-2\pi i n k/N) \\
&= \frac{1}{N} X(0) + \frac{1}{N} \sum_{k=1}^{N-1} X(N-k) \exp(2\pi i n(N-k)/N) \\
&= \frac{1}{N} X(0) + \frac{1}{N} \sum_{k=1}^{N-1} X(k) \exp(2\pi i n k/N) \\
&= x(n).
\end{aligned}$$

This shows that all samples are real numbers and $\mathbf{x}$ is a real-valued vector.

Exercise 2.25. Specify the DFT_N matrix explicitly for $N \in \{1,2,4\}$. Count the number of multiplications and additions when performing the usual matrix–vector product $\mathrm{DFT}_4 \cdot \mathbf{x}$ for a vector $\mathbf{x} = (x_1, x_2, x_3, x_4)^\top$. Then conduct all steps of the FFT algorithm (two recursions are needed) and again count the overall number of multiplications and additions needed to compute $\mathrm{DFT}_4 \cdot \mathbf{x}$.

Solution to Exercise 2.25. For $N = 1$, we obtains $\mathrm{DFT}_1 = (1)$. For $N = 2$, we have $\omega := \exp(-2\pi i/2) = -1$ and

$$\mathrm{DFT}_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}.$$

For $N = 4$, we have $\omega := \exp(-2\pi i/4) = i$ and

$$\mathrm{DFT}_4 = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & i & -1 & -i \\ 1 & 1 & -1 & 1 \\ 1 & -i & -1 & i \end{pmatrix}.$$

When computing the usual matrix–vector product $\mathrm{DFT}_4 \cdot \mathbf{x}$, one needs 16 multiplications and 12 additions.

Applying the FFT algorithm from Table 2.1, the first recursion involves the computation of two DFT_2. Furthemore, to assemble the result, one requires 4 multiplications for the twiddle factors and 4 additions.

In the second recursion, each of DFT_2 involves the computation of two DFT_1, which is free of cost (as DFT_1 of a number is just the number itself). Furthemore, to assemble the result, one requires 2 multiplications for the twiddle factors and 2 additions.

Altogether, using the FFT for DFT_4, one requires $4 + 2 \cdot 2 = 8$ multiplications and $4 + 2 \cdot 2 = 8$ additions.

Exercise 2.26. Let $N = 2^n$ be a power of two. In (2.127), we derived the estimate $\mu(N) \leq 2\mu(N/2) + 1.5N$ for the number of multiplications and additions needed to compute the matrix–vector product $\mathrm{DFT}_N \cdot \mathbf{x}$. Using $\mu(1) = 0$ (the case $n = 0$), show by a mathematical induction on n that this implies $\mu(N) \leq 1.5N \log_2(N)$.

Solution to Exercise 2.26. Suppose that the assertion has been shown for $N = 2^n$ for $n \geq 0$. Then, one obtains for the case $2N = 2^{n+1}$:

$$\begin{aligned} \mu(2N) &\leq 2\mu(N) + 1.5(2N) \\ &\leq 2(1.5N \log_2(N)) + 1.5(2N) \\ &\leq 3N(\log_2(N) + 1) \\ &\leq 1.5(2N) \log_2(2N). \end{aligned}$$

Exercise 2.27. In the spectrograms shown in Figure 2.32 one can notice vertical stripes at $t = 0$ and $t = 1$. Why?

Solution to Exercise 2.27. The signal f is defined in the time interval $[0, 1]$ by (2.142). Furthermore, it is assumed to be zero outside this interval. Now, in a neighborhood of $t = 0$, the signal is zero for $t < 0$ and it is a superposition of two sinusoids for $t > 0$. In the Fourier representation, two exponential functions are needed to represent the signal for $t > 0$. However, for $t < 0$ these oscillations need to compensated

to generate the zero function. To this end, based on the principles of destructive interference, many different frequency components spread over the entire spectrum are needed, which explains the vertical stripe in the spectrogram at $t = 0$. The same explanation applies for $t = 1$.

Exercise 2.28. In this exercise, we prove the **sampling theorem**. A CT-signal $f \in L^2(\mathbb{R})$ is called **Ω-bandlimited** if the Fourier transform $\hat{f}$ vanishes for $|\omega| > \Omega$, i.e., $\hat{f}(\omega) = 0$ for $|\omega| > \Omega$. Let $f \in L^2(\mathbb{R})$ be an Ω-bandlimited function and let x be the T-sampled version of f with $T := 1/(2\Omega)$, i.e., $x(n) = f(nT)$, $n \in \mathbb{Z}$. Then f can be reconstructed from x by

$$f(t) = \sum_{n\in\mathbb{Z}} x(n)\operatorname{sinc}\left(\frac{t-nT}{T}\right) = \sum_{n\in\mathbb{Z}} f\left(\frac{n}{2\Omega}\right)\operatorname{sinc}(2\Omega t - n),$$

where the sinc function is defined in (2.96). In other words, the CT-signal f can be perfectly reconstructed from the DT-signal obtained by equidistant sampling if the bandlimit is no greater than half the sampling rate.
[**Hint:** Note that one may assume $\Omega = 1/2$ (and $T = 1$) by considering the scaled function $t \mapsto f(t/\Omega)$. In this case, f is $1/2$-bandlimited and can be extended to a 1-periodic function g. Represent g by its Fourier series (2.79) and compute the Fourier coefficients $c_n = \langle g|\mathbf{exp}_n\rangle$, $n \in \mathbb{Z}$. Compare these coefficients with the Fourier representation (2.91) of f evaluated at $t = n$ for $n \in \mathbb{Z}$ (again using the fact that f is $1/2$-bandlimited). As a result, one obtains $c_n = f(-n)$. Finally, reconstruct f from the Fourier series of g. To this end, you need the result of Exercise 2.21.]

Solution to Exercise 2.28. Let f be an Ω-bandlimited signal with $\Omega = 1/2$. Then $\hat{f}$ can be extended to a 1-periodic function, which we denote by g. The function g can be represented by its Fourier series (2.79) as

$$g(t) = \sum_{n\in\mathbb{Z}} c_n \exp(2\pi i n t).$$

By (2.80), the coefficients are

$$c_n = \langle g|\mathbf{exp}_n\rangle = \int_{\omega\in[0,1)} g(\omega)\exp(-2\pi i n\omega)d\omega = \int_{|\omega|\le 1/2} g(\omega)\exp(-2\pi i n\omega)d\omega.$$

Next, since f is $(1/2)$-bandlimited, the Fourier representation (2.91) for CT-signals yields:

$$f(t) = \int_{\omega\in\mathbb{R}} \hat{f}(\omega)\exp(2\pi i\omega t)d\omega = \int_{|\omega|\le 1/2} \hat{f}(\omega)\exp(2\pi i\omega t)d\omega$$

and therefore

$$f(-n) = \int_{|\omega|\le 1/2} \hat{f}(\omega)\exp(-2\pi i\omega n)d\omega.$$

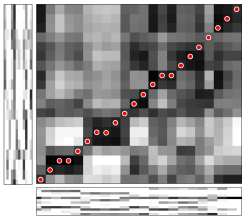
It follows that $c_n = f(-n)$ and therefore

$$g(t) = \sum_{n\in\mathbb{Z}} f(n)\exp(-2\pi int).$$

Using the result of Exercise 2.21, we obtain from this:

$$\begin{aligned} f(t) &= \int_{|\omega|\le 1/2} g(\omega)\exp(2\pi i\omega t)d\omega \\ &= \int_{|\omega|\le 1/2} \sum_{n\in\mathbb{Z}} f(n)e^{-2\pi in\omega}\exp(2\pi i\omega t)d\omega \\ &= \sum_{n\in\mathbb{Z}} f(n) \underbrace{\int_{|\omega|\le 1/2} \exp(2\pi i\omega(t-n))d\omega}_{=\mathrm{sinc}(t-n)}. \end{aligned}$$

Chapter 3
Music Synchronization



Exercise 3.1. In Section 3.1.1, we computed a log-frequency spectrogram based on a semitone resolution using (3.3) and (3.4). In this exercise, we want to specify a log-frequency spectrogram with a resolution of half a semitone (resulting in 24 bands per octave). Write a small computer program that calculates the corresponding center frequencies, the cutoff frequencies, and the bandwidths for the various log-frequency bands, each corresponding to a half semitone (as in Table 3.1). Output all numbers for the resulting 25 bands between C4 and C5. Then, do the same for a log-frequency spectrogram with a resolution of a third semitone (resulting in 36 bands per octave). Again, output all numbers for the resulting 37 bands between C4 and C5.

Solution to Exercise 3.1.

```
SemitoneRes = 1/2;
for p = 60:SemitoneRes:72
    CF = 2^((p-69)/12)*440;
    CutL = 2^((p-SemitoneRes/2-69)/12)*440;
    CutU = 2^((p+SemitoneRes/2-69)/12)*440;
    BW = CutU - CutL;
    fprintf('p = %4.2f, ',p);
    fprintf('CF = %6.2f, ',CF);
    fprintf('CutL = %6.2f, ',CutL);
    fprintf('CutU = %6.2f, ',CutU);
    fprintf('BW = %4.2f\n',BW);
end

p = 60.00, CF = 261.63, CutL = 257.87, CutU = 265.43, BW = 7.56
p = 60.50, CF = 269.29, CutL = 265.43, CutU = 273.21, BW = 7.78
p = 61.00, CF = 277.18, CutL = 273.21, CutU = 281.21, BW = 8.01
...

SemitoneRes = 1/3;
...
p = 60.00, CF = 261.63, CutL = 259.12, CutU = 264.16, BW = 5.04
p = 60.33, CF = 266.71, CutL = 264.16, CutU = 269.29, BW = 5.14
p = 60.67, CF = 271.90, CutL = 269.29, CutU = 274.53, BW = 5.24
```

Exercise 3.2. Assuming a sampling rate of $F_s = 44100$ Hz and a window length of $N = 4096$, determine the largest pitch p for which the set $P(p)$ defined in (3.3) is empty. What are the center frequency, the cutoff frequencies, and the bandwidth of the corresponding log-frequency band?

Solution to Exercise 3.2. For $p = 51$, one obtains, $F_{\text{pitch}}(p) = 155.56$ Hz, $F_{\text{pitch}}(p - 0.5) = 151.13$ Hz, $F_{\text{pitch}}(p + 0.5) = 160.12$ Hz, and $\text{BW}(p) = 8.99$. Furthermore, by (2.28), we have $F_{\text{coef}}(k) = (k \cdot F_s)/N = k \cdot 44100/4096 \approx k \cdot 10.77$ Hz. From this, one obtains $F_{\text{coef}}(15) = 150.73$ Hz and $F_{\text{coef}}(16) = 161.50$ Hz, which shows that $P(p)$ is empty for $p = 51$. To show that $p = 51$ is largest p with this property, one checks that $\text{BW}(p)$ is nonempty for $p = \{52, 53, 54\}$. Furthermore, for $p \geq 55$ one obtains $\text{BW}(p) > F_s/N$, which shows that $\text{BW}(p)$ is nonempty for $p \geq 55$.

Exercise 3.3. Let $\text{BW}(p) = F_{\text{pitch}}(p + 0.5) - F_{\text{pitch}}(p - 0.5)$ be the bandwith for a pitch p as defined in (3.5). What is the relation between the bandwidths $\text{BW}(p+12)$ and $\text{BW}(p)$ of two pitches that are one octave apart? Give a mathematical proof for your claim. Similarly, determine the relation between the bandwidths $\text{BW}(p + 1)$ and $\text{BW}(p)$ of two neighboring pitches.

Solution to Exercise 3.3. Using (3.2), we obtain

$$F_{\text{pitch}}(r + 12) = 2^{(r+12-69)/12} \cdot 440 = 2 \cdot F_{\text{pitch}}(r)$$

for any $r \in \mathbb{R}$. Applying this to $r = p - 0.5$ and $r = p + 0.5$, we obtain from (3.5):

$$\begin{aligned} \text{BW}(p + 12) &= F_{\text{pitch}}(p + 12 + 0.5) - F_{\text{pitch}}(p + 12 - 0.5) \\ &= 2 \cdot F_{\text{pitch}}(p + 0.5) - 2 \cdot F_{\text{pitch}}(p + 0.5) \\ &= 2 \cdot \text{BW}(p). \end{aligned}$$

In other words, increasing the pitch by one octave increases the bandwith by a factor of two. Similarly, one shows that $\text{BW}(p + 1)/\text{BW}(p) = 2^{1/12}$ (see also Exercise 1.6).

Exercise 3.4. Given an audio signal at a sampling rate of $F_s = 22050$ Hz, we want to compute a log-frequency spectrogram as in (3.4). As a requirement, all sets $P(p)$ (as defined in (3.3)) for all pitches corresponding to the notes C2 ($p = 36$) to C3 ($p = 48$) should contain at least four Fourier coefficients. To meet this requirement, what is the minimal window length N (assuming that N is a power of two) to be used in the STFT? For this N, determine the elements of the set $P(36)$ explicitly.

Solution to Exercise 3.4. For pitch $p = 36$ (corresponding to C2), one obtains $F_{\text{pitch}}(p) = 65.41$ Hz, $F_{\text{pitch}}(p - 0.5) = 63.54$ Hz, $F_{\text{pitch}}(p + 0.5) = 67.32$ Hz, and $\text{BW}(p) = 3.78$. Furthermore, for $F_s = 22050$ Hz and $N = 32768$, it follows from (2.28) that $F_{\text{coef}}(96) = 63.93$ Hz, $F_{\text{coef}}(101) = 67.29$ Hz, and $P(36) = \{96, \ldots, 101\}$. In other words, $P(36)$ contains six elements for $N = 32768$. This also shows, that $P(p)$ must contain at least four elements for $p > 36$. Finally, for the window size $N = 16384$, the set $P(36)$ contains only three elements. Therefore, $N = 32768$ is the minimal window length with the desired property.

Exercise 3.5. The tuning of musical instruments is usually based on a fixed reference pitch. In Western music, one typically uses the **concert pitch** A4 having a frequency of 440 Hz (see Section 1.3.2). To estimate the deviation from this ideal reference, a musician is asked to play the note A4 on his or her instrument over the duration of four seconds. Describe a simple FFT-based procedure for estimating the tuning deviation of the instrument used. How would you choose the parameters (sampling rate, window size) to obtain an accuracy of at least 1 Hz in this estimation?

Solution to Exercise 3.5. Playing a note A4, one can expect dominant frequencies in a neighborhood of 440 Hz (corresponding to the fundamental frequency) and its integer multiples (corresponding to the harmonics). One basic procedure is to first compute a DFT of the recorded signal to obtain a spectral representation. Then, one may look for the frequency index k_0 that yields a maximal magnitude coefficient $|F_{\mathrm{coef}}(k_0)|$ in a neighborhood of 440 Hz (e.g., plus/minus a semitone). The difference $F_{\mathrm{coef}}(k_0) - 440$ Hz then yields an estimate of the tuning deviation.

Assuming a sampling rate of $F_{\mathrm{s}} = 44100$ Hz, one may compute a DFT using a window size of $N = 2^{17} = 131072$ (corresponding to 2.97 sec). This yields a spectral resolution of $F_{\mathrm{s}}/N \approx 0.34$ Hz.

To obtain a more robust estimate, one may also consider spectral peak positions in suitably defined neighborhoods of the harmonics. The resulting deviations from the ideal positions of the individual harmonics can be used to derive a single tuning estimate using a suitable fusion strategy.

Exercise 3.6. Assume that an orchestra is tuned 20 cents upwards compared with the standard tuning. What is the center frequency of the tone A4 in this tuning? How can a chroma representation be adjusted to compensate for this tuning difference?

Solution to Exercise 3.6. The detuning of 20 cents corresponds to a fifth of a semitone. Therefore, by (3.2), the center frequency of the tone A4 in the given tuning is

$$F_{\mathrm{pitch}}(69.2) = 2^{(69.2-69)/12} \cdot 440 \approx 445.1 \text{ Hz}.$$

Using this frequency as a new reference, we define the function

$$F'_{\mathrm{pitch}}(p) = 2^{(p-69)/12} \cdot 445.1.$$

Based on this modified function, we define a modified set

$$P'(p) := \{k : F'_{\mathrm{pitch}}(p-0.5) \leq F_{\mathrm{coef}}(k) < F'_{\mathrm{pitch}}(p+0.5)\}$$

for each pitch $p \in [0:127]$ (see 3.3). From this, we obtain an adjusted log-frequency spectrogram (see (3.4)), from which we can derive an adjusted chroma representation as before (see (3.6)).

Exercise 3.7. Show that the DTW distance as defined in (3.21) is symmetric (i.e., $\mathrm{DTW}(X,Y) = \mathrm{DTW}(Y,X)$ for any two given sequences $X = (x_1, x_2, \ldots, x_N)$ and $Y = (y_1, y_2, \ldots, y_M)$) in the case that the local cost measure c is symmetric.

Solution to Exercise 3.7. Let $P = (p_1, \ldots, p_L)$ with $p_\ell = (n_\ell, m_\ell) \in [1:N] \times [1:M]$ for $\ell \in [1:L]$ be an (N,M)-warping path. Setting $p'_\ell := (m_\ell, n_\ell)$, one easily checks that the conditions (3.16), (3.17), and (3.18) are satisfied for $P' := (p'_1, \ldots, p'_L)$. In other words, P' defines an (M,N)-warping path. By this assignment, one obtains a one-to-one correspondence between the set of (N,M)-warping paths and the set of (M,N)-warping paths. Furthermore, in the case that c is symmetric, one obtains

$$c_{P'}(Y,X) = \sum_{\ell=1}^{L} c(y_{m_\ell}, x_{n_\ell}) = \sum_{\ell=1}^{L} c(x_{n_\ell}, y_{m_\ell}) = c_P(X,Y).$$

Therefore, an (M,N)-warping path P' has minimal cost if and only if the corresponding (N,M)-warping path P has minimal cost. From this and (3.21), we obtain $\mathrm{DTW}(Y,X) = \mathrm{DTW}(X,Y)$, which proves the symmetry of the DTW distance.

Exercise 3.8. Let $P = (p_1, p_2, \ldots p_L)$ be an arbitrary (N,M)-warping path. Specify the smallest possible lower bound as well as the largest possible upper bound for the length L of P in terms of N and M.

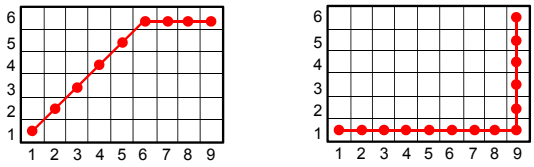
Solution to Exercise 3.8. Because of the boundary condition (3.16) and the step size condition (3.18), each element of the first sequence must be assigned to an element of the second sequence, which implies $L \geq N$. Similarly, each element of the second sequence must be assigned to an element of the first sequence, which implies $L \geq M$. This proves

$$\max(N,M) \leq L.$$

Because of the monotonicity condition (3.17), a warping path must increase in either the first or the second dimension (or both). This implies

$$L \leq N + M - 1.$$

The following examples indicate that the specified upper and lower bounds may be assumed by warping paths, which shows that the bounds are optimal.



Exercise 3.9. In this exercise, we show that there is a large number of theoretically possible warping paths. Let $\mu(N,M)$ be the number of possible (N,M)-warping paths for some given N and M. Obviously, in the case $N = 1$ or $M = 1$, there is only one possible warping path, i.e., $\mu(1,M) = \mu(N,1) = 1$. Show that $\mu(2,2) = 3$, $\mu(2,3) = 5$, and $\mu(3,3) = 13$. Derive a general recursive formula for $\mu(N,M)$ for $N > 1$ and $M > 1$. Compute $\mu(N,M)$ for $(N,M) \in [1:6] \times [1:6]$.

Solution to Exercise 3.9. Let $N > 1$ and $M > 1$. To reach the cell (N,M) by a warping path, one has three possibilities: either one comes from $(N-1,M)$, or from

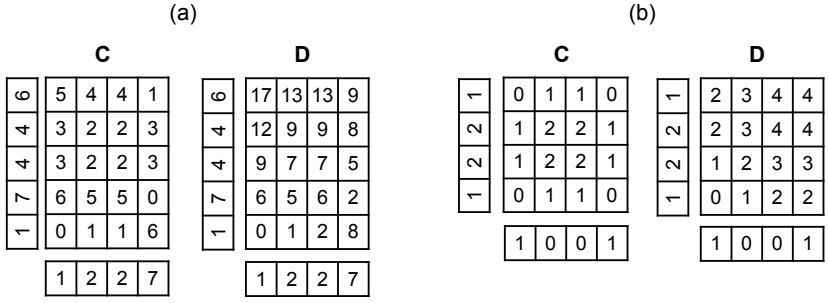
$(N,M-1)$, or from $(N-1,M-1)$. Therefore, $\mu(N,M) = \mu(N-1,M) + \mu(N,M-1) + \mu(N-1,M-1)$ for $N > 1$ and $M > 1$. From this, one obtains the following values $\mu(N,M)$ for $(N,M) \in [1:6] \times [1:6]$:

$$\begin{bmatrix} 1 & 11 & 61 & 231 & 681 & 1683 \\ 1 & 9 & 41 & 129 & 321 & 681 \\ 1 & 7 & 25 & 63 & 129 & 231 \\ 1 & 5 & 13 & 25 & 41 & 61 \\ 1 & 3 & 5 & 7 & 9 & 11 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Exercise 3.10. Let $F = \mathbb{R}$ be a feature space and $c : \mathcal{F} \times \mathcal{F} \to \mathbb{R}_{\geq 0}$ a local cost measure defined by $c(x,y) = |x-y|$ for $x,y \in \mathbb{R}$. Compute $\mathrm{DTW}(X,Y)$ for the following sequences X and Y as well as all optimal warping paths. Also specify the cost matrix **C** and the accumulated cost matrix **D**.

(a) $X = (1,7,4,4,6)$ and $Y = (1,2,2,7)$.
(b) $X = (1,2,2,1)$ and $Y = (1,0,0,1)$.

Solution to Exercise 3.10. The cost matrix **C** and the accumulated cost matrix **D** are as follows:



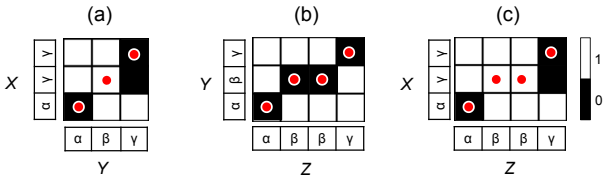
For (a), one obtains $\mathrm{DTW}(X,Y) = 9$. The only optimal warping is given by $((1,1),(1,2),(1,3),(2,4),(3,4),(4,4),(5,4))$. For (b), one obtains $\mathrm{DTW}(X,Y) = 4$. There exist the following seven optimal warping paths:

1.) $((1,1),(2,1),(3,1),(4,1),(4,2),(4,3),(4,4))$.
2.) $((1,1),(2,1),(3,1),(4,2),(4,3),(4,4))$.
3.) $((1,1),(2,1),(3,2),(4,3),(4,4))$.
4.) $((1,1),(2,2),(3,3),(4,4))$.
5.) $((1,1),(1,2),(1,3),(1,4),(2,4),(3,4),(4,4))$.
6.) $((1,1),(1,2),(1,3),(2,4),(3,4),(4,4))$.
7.) $((1,1),(1,2),(2,3),(3,4),(4,4))$.

Exercise 3.11. In this excercise, we show that the DTW distance generally does not satisfy the triangle inequality. Let $\mathcal{F} := \{\alpha,\beta,\gamma\}$ be an abstract feature space consisting of three different elements. Define a cost measure $c : \mathcal{F} \times \mathcal{F} \to \{0,1\}$

by setting $c(x,y) := 1 - \delta_{xy}$ for $x,y \in \mathcal{F}$. In other words, $c(x,y) := 0$ if $x = y$ and $c(x,y) := 1$ if $x \neq y$ for $x,y \in \mathcal{F}$. Note that c defines a metric on $\mathcal{F}$ and, in particular, satisfies the triangle inequality. Now, consider the three sequences $X := (\alpha,\gamma,\gamma)$, $Y := (\alpha,\beta,\gamma)$, and $Z := (\alpha,\beta,\beta,\gamma)$ over $\mathcal{F}$. Compute $\mathrm{DTW}(X,Y)$, $\mathrm{DTW}(Y,Z)$, and $\mathrm{DTW}(X,Z)$. Furthermore, show that the triangle inequality does not hold in this example.

Solution to Exercise 3.11. The following figure indicates the cost matrices and optimal warping paths:



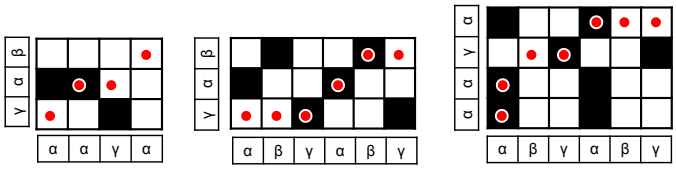
This yields $\mathrm{DTW}(X,Y) = 1$, $\mathrm{DTW}(Y,Z) = 0$, and $\mathrm{DTW}(X,Z) = 2$. Furthermore, one has

$$\mathrm{DTW}(X,Z) = 2 > 1 = \mathrm{DTW}(X,Y) + \mathrm{DTW}(Y,Z),$$

which shows that the triangle inequality does not hold in this example.

Exercise 3.12. Let $\mathcal{F} = \{\alpha,\beta,\gamma\}$ and $c : \mathcal{F} \times \mathcal{F} \to \{0,1\}$ be as in Exercise 3.11. Specify the DTW distances $\mathrm{DTW}(X,Y)$, $\mathrm{DTW}(X,Z)$, and $\mathrm{DTW}(Y,Z)$ for the sequences $X = (\gamma,\alpha,\beta)$, $Y = (\alpha,\alpha,\gamma,\alpha)$, and $Z = (\alpha,\beta,\gamma,\alpha,\beta,\gamma)$. Instead of using the dynamic programming approach, try to "guess" the DTW distances by specifying suitable warping paths. Then, argue that the specified warping paths are indeed optimal.

Solution to Exercise 3.12. The following figure specifies warping paths all of which having a total cost of three:



We now need to argue that there are no warping paths of lower total cost.

As for X and Y, the boundary condition implies that $x_1 = \gamma$ needs to be aligned to $y_1 = \alpha$ and $x_3 = \beta$ to $y_4 = \alpha$, which already leads to a cost of two. Furthermore, $y_3 = \gamma$ is either aligned to $x_1 = \gamma$ without cost, but then $y_2 = \alpha$ needs also to be aligned to $x_1 = \gamma$ because of the monotonicity condition, thus resulting in another cost of one. Or $y_3 = \gamma$ is aligned to x_2 or x_3, both resulting in a cost of one. Altogether we have shown that *any* warping path has at least a total cost of three.

As for X and Z, the boundary condition again implies a cost of two. Furthermore, $z_2 = \beta$ is either aligned to x_3 without cost, which then implies that $z_3 = \gamma$ needs also

to be aligned to x_3. Or $z_2 = \beta$ is aligned to x_2 or x_3, both being a mismatch resulting in a cost of one. Again the total cost is at least three.

As for X and Z, the boundary condition implies a cost of one. Since Z contains β two times and Y does not contain this element at all, this results in an additional cost of two. Again the total cost is at least three.

Altogether, we have shown that $\mathrm{DTW}(X,Y) = 3$, $\mathrm{DTW}(X,Z) = 3$, and $\mathrm{DTW}(Y,Z) = 3$.

Exercise 3.13. Extend the accumulated cost matrix $\mathbf{D}$ from Section 3.2.1.3 by an additional row and column indexed by 0. Define $\mathbf{D}(n,0) := \infty$ for $n \in [1:N]$, $\mathbf{D}(0,m) := \infty$ for $m \in [1:M]$, and $\mathbf{D}(0,0) := 0$. Show that one obtains the original accumulated cost matrix when applying the recursion of (3.25) for $n \in [1:N]$ and $m \in [1:M]$.
[**Hint:** When computing with the value ∞, we assume that the sum of the value ∞ with a finite value is defined to be ∞. Furthermore, the minimum over a set containing finite values as well as the value ∞ is defined to be the minimum over the finite values.]

Solution to Exercise 3.13. Let us first consider the case $n = 1$ and $m = 1$. Applying (3.25), one obtains

$$\begin{aligned}\mathbf{D}(1,1) &= \mathbf{C}(1,1) + \min\{\mathbf{D}(0,0), \mathbf{D}(0,1), \mathbf{D}(1,0)\} \\ &= \mathbf{C}(1,1) + \min\{0, \infty, \infty\} \\ &= \mathbf{C}(1,1).\end{aligned}$$

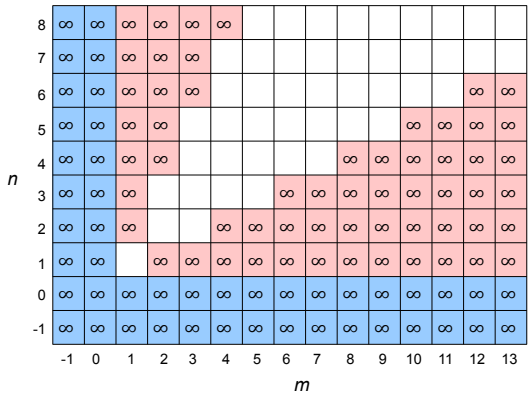
Next, for $n > 1$ and $m = 1$, one obtains

$$\begin{aligned}\mathbf{D}(n,1) &= \mathbf{C}(n,1) + \min\{\mathbf{D}(n-1,0), \mathbf{D}(n-1,1), \mathbf{D}(n,0)\} \\ &= \mathbf{C}(n,1) + \min\{\infty, \mathbf{D}(n-1,1), \infty\} \\ &= \mathbf{C}(n,1) + \mathbf{D}(n-1,1).\end{aligned}$$

From this and $\mathbf{D}(1,1) = \mathbf{C}(1,1)$, one obtains $\mathbf{D}(n,1) = \sum_{k=1}^{n} \mathbf{C}(k,1)$ for $n \in [1:N]$, which is (3.23). Similarly, one shows $\mathbf{D}(1,m) = \sum_{k=1}^{m} \mathbf{C}(1,k)$ for $m \in [1:M]$, which is (3.24). In other words, one obtains the same initialization as for the original approach. As a consequence, the recursion (3.25) for $n \in [2:N]$ and $m \in [2:M]$ yields the same values as the original approach.

Exercise 3.14. In this exercise, we consider DTW with the step size condition $\Sigma = \{(2,1), (1,2), (1,1)\}$ (see (3.30)). As in Exercise 3.13, we extend the accumulated cost matrix $\mathbf{D}$, this time by two additional rows and columns indexed by -1 and 0. Then we set $\mathbf{D}(1,1) := \mathbf{C}(1,1)$, $\mathbf{D}(n,-1) := \mathbf{D}(n,0) := \infty$ for $n \in [-1:N]$, and $\mathbf{D}(-1,m) := \mathbf{D}(0,m) := \infty$ for $m \in [-1:M]$. $\mathbf{D}$ is then computed using the recursion of (3.31) for $n \in [1:N]$ and $m \in [1:M]$. Specify the cells $(n,m) \in [-1:N] \times [-1:M]$ for which one obtains $\mathbf{D}(n,m) = \infty$. Furthermore, describe some meaningful constraints for the lengths N and M in this alignment scenario.

Solution to Exercise 3.14 . By definition $\mathbf{D}(1,1) = \mathbf{C}(1,1) < \infty$. For the cells $(n,1)$ for $n \in [2:N]$ and $(1,m)$ for $m \in [2:M]$, the recursion (3.31) yields $\mathbf{D}(n,1) = \mathbf{D}(1,m) = \infty$. For general $n \in [1:N]$ and $1 \in [1:M]$, as illustrated by the following figure, one obtains $\mathbf{D}(n,m) = \infty$ if $(n/m) \leq 1/2$ or $(n/m) \geq 2$.



As a consequence, when $N \geq 2M$ or $M \geq 2N$, all cells of $\mathbf{D}$ have a value of ∞, which reflects the fact that no alignment is possible in this case using the step size condition $\Sigma = \{(2,1),(1,2),(1,1)\}$. Therefore, in this alignment scenario, one needs to enforce $1/2 < N/M < 2$.

Exercise 3.15. Let $F = \mathbb{R}$ be a feature space and $c : \mathcal{F} \times \mathcal{F} \to \mathbb{R}_{\geq 0}$ a local cost measure defined by $c(x,y) = |x-y|$ for $x,y \in \mathbb{R}$ (see Exercise 3.10). Compute $\mathrm{DTW}(X,Y)$ for the sequences $X = (1,7,4,4,6)$ and $Y = (1,2,2,7)$ as well as all optimal warping paths using the step size condition $\Sigma = \{(2,1),(1,2),(1,1)\}$ from (3.30). Also specify the cost matrix $\mathbf{C}$ and the accumulated cost matrix $\mathbf{D}$ using two additional rows and columns initialized with ∞ (see Exercise 3.14).

Solution to Exercise 3.15.

C:

6	5	4	4	1
4	3	2	2	3
4	3	2	2	3
7	6	5	5	0
1	0	1	1	6
	1	2	2	7

D:

6	∞	∞	∞	∞	6	5
4	∞	∞	∞	∞	4	5
4	∞	∞	∞	2	7	8
7	∞	∞	∞	5	5	∞
1	∞	∞	0	∞	∞	∞
	∞	∞	∞	∞	∞	∞
	∞	∞	∞	∞	∞	∞
	1	2	2	7		

Therefore, $\mathrm{DTW}(X,Y) = 5$. The only optimal warping is given by $((1,1),(3,2),(4,3),(5,4))$.

Exercise 3.16. In software such as MATLAB, an operation expressed as a matrix product can often be computed more efficiently than, e.g., using nested loops over the matrix indices. This motivates the following exercise. Let $c : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ be

the cosine distance for $\mathcal{F} = \mathbb{R}^{12} \setminus \{0\}$ (see (3.14)). Given two feature sequences $X = (x_1, x_2, \ldots, x_N)$ and $Y = (y_1, y_2, \ldots, y_M)$ over $\mathcal{F}$, let $\mathbf{C}(n,m) := c(x_n, y_m)$ be the resulting cost matrix for $n \in [1:N]$ and $m \in [1:M]$ (see (3.13)). Show how $\mathbf{C}$ can be computed using matrix products (instead of a nested loop over the indices n and m to compute the individual entries $\mathbf{C}(n,m)$).

Solution to Exercise 3.16 . Let $D = 12$ be the dimension of the feature space $\mathcal{F}$. We construct an $(N \times D)$ matrix $\mathbf{X}$ by taking the normalized vectors $x_n/\|x_n\|$ as columns. Similarly, we construct an $(M \times D$ matrix $\mathbf{Y}$ by taking the normalized vectors $y_m/\|y_m\|$ as columns. Let $\mathbf{Y}^\top$ denoted the transposed matrix of $\mathbf{Y}$. Then the matrix product

$$\mathbf{X} \cdot \mathbf{Y}^\top$$

defines an $(N \times M)$ matrix. Let $\mathbf{1}$ be the all-ones matrix of dimension $N \times M$. Then we obtain

$$\begin{aligned}
(\mathbf{1} - \mathbf{X} \cdot \mathbf{Y}^\top)(n,m) &= \mathbf{1}(n,m) - \sum_{d=1}^{D} \mathbf{X}(n,d)\mathbf{Y}^\top(d,m) \\
&= 1 - \sum_{d=1}^{D} \frac{x_n(d)}{\|x_n\|} \frac{y_m(d)}{\|y_m\|} \\
&= 1 - \frac{\langle x_n | y_m \rangle}{\|x_n\| \|y_m\|} \\
&= \mathbf{C}(n,m)
\end{aligned}$$

In other words, $\mathbf{C} = \mathbf{1} - \mathbf{X} \cdot \mathbf{Y}^\top$.

Exercise 3.17. Assume that, for two given sequences $X = (x_1, \ldots, x_N)$ and $Y = (y_1, \ldots, y_M)$, there is exactly one optimal (N,M)-warping path denoted by P^*. Furthermore, let $R \subseteq [1:N] \times [1:M]$ be a global constraint region (see Section 3.2.2.3). Show that the constrained optimal warping path P_R^* coincides with P^* if and only if P^* is contained in R.

Solution to Exercise 3.17. If P^* is not contained in R, then it is clear that P_R^* cannot coincide with P^*. Next, let us assume that P^* is contained in R. Since $c_{P^*}(X,Y)$ has minimal cost over all possible (N,M)-warping paths in $[1:N] \times [1:M]$, it must also have minimal cost over all possible (N,M)-warping paths that lie in the constraint region $R \subseteq [1:N] \times [1:M]$. Therefore, $P_R^* = P^*$.

Exercise 3.18. In this exercise, we analyze the multiscale approach to DTW (MsDTW) as outlined in Section 3.2.2.4. Let $X = (x_1, x_2, \ldots, x_N)$ and $Y = (y_1, y_2, \ldots, y_M)$ be sequences of length N and M, respectively. For simplicity, we assume that $N = M = 2^K$ for a natural number $K \in \mathbb{N}$. Let $A^{\mathrm{DTW}}(N) = N^2$ denote the number of evaluations of the local cost measure that are required in the classical DTW algorithm. Furthermore, we assume that we have a coarsening and downsampling procedure for computing the coarsened sequences $X_1, X_2, \ldots, X_K$ and $Y_1, Y_2, \ldots, Y_K$, where the sampling rates are successively reduced by factors

$f_1 = f_2 = \ldots = f_K = 2$. In the subsequent analysis, we neglect the operations required for the coarsening and downsampling procedure. Let $A^{\mathrm{MsDTW}}(N)$ denote the number of evaluations of the local cost measure that are required in the MsDTW algorithm. Specify a recursive equation for $A^{\mathrm{MsDTW}}(N)$. Derive from this equation an upper bound for $A^{\mathrm{MsDTW}}(N)$.
[**Hint:** Look at an upper bound for the length of a warping path at level k, $1 \leq k \leq K$.]

Solution to Exercise 3.18. At level k, $1 \leq k \leq K$, both of the sequences X_k and Y_k have length 2^{K-k+1}. Let L_k be the length of the warping path between X_k and Y_k, then

$$L_k \leq 2 \cdot 2^{K-k+1} = 2^{K-k+2}.$$

In other words, $L_1 \leq 2N$, $L_2 \leq 2(N/2) = N$, and so on. Then the following holds:

$$\begin{aligned}
A^{\mathrm{MsDTW}}(N) &= A^{\mathrm{MsDTW}}\left(\frac{N}{2}\right) + f_1^2 \cdot L_2 \\
&\leq A^{\mathrm{MsDTW}}\left(\frac{N}{2}\right) + 4 \cdot N \\
&= A^{\mathrm{MsDTW}}\left(\frac{N}{2^2}\right) + f_2^2 \cdot L_3 + 4 \cdot N \\
&\leq A^{\mathrm{MsDTW}}\left(\frac{N}{2^2}\right) + 4 \cdot \left(\frac{N}{2}\right) + 4 \cdot N \\
&\leq \ldots \\
&\leq A^{\mathrm{MsDTW}}\left(\frac{N}{2^{K-1}}\right) + 4 \cdot \left(\frac{N}{2^{K-2}}\right) \ldots + 4 \cdot \left(\frac{N}{2^2}\right) + 4 \cdot \left(\frac{N}{2}\right) + 4 \cdot N \\
&\leq A^{\mathrm{DTW}}(2) + 4 \cdot \left(4 + \ldots + 2^{K-2} + 2^{K-1} + 2^K\right) \\
&\leq 4 \cdot \sum_{k=0}^{K} 2^k \\
&\leq 4 \cdot 2^{K+1} \\
&= 8N
\end{aligned}$$

Exercise 3.19. In computer science, the **edit distance** (sometimes also referred to as the **Levenshtein distance**) is a string metric for measuring the difference between two sequences $X = (x_1, x_2, \ldots, x_N)$ and $Y = (y_1, y_2, \ldots, y_M)$ over an alphabet $\mathcal{F}$. The sequences are also often called **words**, and the elements of the alphabet are called **characters**. The edit distance $\mathrm{Edit}(X,Y)$ between X and Y is defined to be the minimum number of single-character edits required to change one sequence into the other. One allows three kinds of single-character edits referred to as **insertion** (including an additional character), **deletion** (omitting a character of a word), and **substitution** (replacing a character of a word by another character). Develop an algorithm based on dynamic programming (as in Table 3.2) that computes the edit distance between two given sequences X and Y.
[**Hint:** Define an accumulated cost matrix using (3.22). Let ε denote the empty word

of length zero. Use this empty word as a recursion start to compute the accumulated cost matrix. For an example application, see Exercise 3.20.]

Solution to Exercise 3.19. Let $X(1:n) := (x_1, \ldots x_n)$ denote the prefix of length $n \in [0:N]$ of X and $Y(1:m) := (y_1, \ldots y_m)$ the one of length $m \in [0:M]$ of Y. In the cases $n = 0$ and $m = 0$, the prefix consists of the empty word ε. Building upon (3.22), we define the **accumulated cost matrix** by setting

$$\mathbf{D}(n,m) := \mathrm{Edit}(X(1:n), Y(1:m))$$

for $n \in [0:N]$ and $m \in [0:M]$. By definition, one has $\mathrm{Edit}(X,Y) = \mathbf{D}(N,M)$. Furthermore, one obviously has $\mathbf{D}(n,0) = n$ for $n \in [0:N]$ and $\mathbf{D}(0,m) = m$ for $m \in [0:M]$. The remaining values of $\mathbf{D}$ can be computed by the following recursion:

$$\mathbf{D}(n,m) = \min \begin{cases} \mathbf{D}(n-1,m-1) + 1 - \delta(x_n, y_m) \\ \mathbf{D}(n-1,m) + 1 \\ \mathbf{D}(n,m-1) + 1 \end{cases}$$

for $n \in [1:N]$ and $m \in [1:M]$, where $\delta(x_n, y_m)$ assumes the value one if $x_n = y_m$ and the value zero if $x_n \neq y_m$. In this recursion, the first case accounts for a substitution, the second case for a deletion, and the third case for an insertion. The sequence of edits that are applied to change one sequence into the other are obtained by applying a backtracking procedure, similar to the construction of an optimal warping path in the case of DTW (see Table 3.2).

Exercise 3.20. The edit distance as introduced in Exercise 3.19 finds applications in biochemistry to compare the primary structures of biological molecules. In this exercise, we consider the case of **deoxyribonucleic acid** or DNA, which is a molecule that encodes the genetic instructions used in the development and functioning of living organisms. The primary structure of DNA can be specified by a sequence of simpler units called **nucleotides**, which are associated to base components referred to as adenine (A), cytosine (C), guanine (G), and thymine (T). Therefore, the primary structure of a DNA molecule can be specified by a sequence over the alphabet $\mathcal{F} := \Sigma := \{\mathtt{A}, \mathtt{C}, \mathtt{G}, \mathtt{T}\}$. In evolutionary biology, **homology** is the similarity between attributes of organisms (e.g., genes) that results from their shared ancestry. In genetics, homology is measured by comparing DNA sequences. A high sequence similarity between two DNA sequences is an indicator for a high probability of being homologous (e.g., sharing a common ancestor). Typical differences of homologous sequences caused by **mutation** are **substitutions** (e.g., TGAT $\rightsquigarrow$ G̲GAT), **insertions** (e.g., TGAT $\rightsquigarrow$ TĊGAT), and **deletions** (e.g., TGAT $\rightsquigarrow$ T~~G~~AT). This illustrates why the edit distance is suitable for comparing the distance (or similarity) of DNA sequences.

By applying Exercise 3.19, compute the edit distance, the accumulated cost matrix, as well as the sequence of edits for the two sequences $X = \mathtt{TGAT}$ and $Y = \mathtt{CGAGT}$.

Solution to Exercise 3.20. Let ε denote the empty word. Then $\mathbf{D}$ is given by the the following table:

	ε	C	G	A	G	T
ε	**0**	1	2	3	4	5
T	1	**1**	2	3	4	4
G	2	2	**1**	2	3	4
A	3	3	2	**1**	**2**	3
T	4	4	3	2	2	**2**

Therefore, $\mathrm{Edit}(X,Y) = \mathbf{D}(4,5) = 2$. As for the edits, there is one optimal sequence indicated by the bold entries in the table. This information is obtained by backtracking the minimizing cells starting with $(N,M) = (4,5)$. As a result, $X =$ TGAT is transformed by replacing the first character (T $\rightsquigarrow$ C) and then inserting a character (G before the last character): TGAT $\rightsquigarrow$ C̲GAĠT.

Exercise 3.21. Another problem related to DTW and the edit distance is known as the **longest common subsequence** (LCS) problem. Given two sequences $X = (x_1, x_2, \ldots, x_N)$ and $Y = (y_1, y_2, \ldots, y_M)$ over an alphabet $\mathcal{F}$, the goal is to find a longest subsequence common to both sequences. For example, the sequences $X = (b,a,b,c,b)$ and $Y = (a,b,b,c,c,b)$ over the alphabet $\mathcal{F} = \{a,b,c\}$ have the longest common subsequence (a,b,c,b). Develop an algorithm based on dynamic programming (as in Table 3.2) for determining the length $\mathrm{LCS}(X,Y)$ of a longest common subsequence of X and Y. Then, determine a longest common subsequence via backtracking. Finally, apply the algorithm to the two sequences $X = (b,a,b,c,b)$ and $Y = (a,b,b,c,c,b)$.
[**Hint:** Define an accumulated similarity matrix as in (3.22). Let ε denote the empty sequence of length zero. Use this empty sequence as a recursion start for computing the accumulated similarity matrix.]

Solution to Exercise 3.21. As before, let $X(1:n) := (x_1, \ldots x_n)$ be the prefix of length $n \in [0:N]$ of X and $Y(1:m) := (y_1, \ldots y_m)$ the prefix of length $m \in [0:M]$ of Y. In the cases $n = 0$ and $m = 0$, the prefix consists of the empty sequence denoted by ε. As in (3.22), we define an **accumulated similarity matrix** by setting

$$\mathbf{D}(n,m) := \mathrm{LCS}(X(1:n), Y(1:m))$$

for $n \in [0:N]$ and $m \in [0:M]$. By definition, one has $\mathrm{LCS}(X,Y) = \mathbf{D}(N,M)$. Furthermore, one obviously has $\mathbf{D}(n,0) = 0$ for $n \in [0:N]$ and $\mathbf{D}(0,m) = 0$ for $m \in [0:M]$. The remaining values of $\mathbf{D}$ can be computed by the following recursion:

$$\mathbf{D}(n,m) = \max \begin{cases} \mathbf{D}(n-1,m-1) + \delta(x_n, y_m) \\ \mathbf{D}(n-1,m) \\ \mathbf{D}(n,m-1) \end{cases}$$

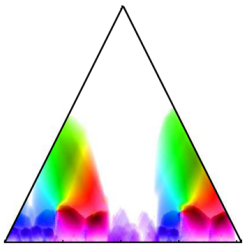
for $n \in [1:N]$ and $m \in [1:M]$, where $\delta(x_n, y_m)$ assumes the value one if $x_n = y_m$ and the value zero if $x_n \neq y_m$. The longest common subsequence can be obtained by a backtracking procedure, similar to the construction of an optimal warping path in the case of DTW (see Table 3.2). In this backtracking, a new common character is found whenever a diagonal step has been performed.

As for the example sequences, the accumulated similarity matrix **D** is given by the following table:

	ε	a	b	b	c	c	b
ε	**0**	0	0	0	0	0	0
b	0	0	1	1	1	1	1
a	0	**1**	1	1	1	1	1
b	0	1	**2**	2	2	2	2
c	0	1	2	2	3	**3**	3
b	0	1	2	3	3	3	**4**

The longest common subsequence is indicated by bold values, which have been obtained by performing a diagonal step. Note that, as in this example, the backtracking may not yield a unique solution. In general, there may be several optimal solutions.

Chapter 4
Music Structure Analysis



Exercise 4.1. Let $\mathcal{F} = \mathbb{R}^D$ be the real vector space of dimension $D \in \mathbb{N}$. Typical similarity measures are based on the Euclidean norm (also referred to as the ℓ^2-norm) defined by

$$\|x\|_2 := \left(\sum_{i=1}^{D} |x(i)|^2\right)^{1/2}$$

for a vector $x = (x(1), x(2), \ldots, x(D))^\top$. From this norm, one can derive the similarity measures $s^{a,b} : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ for constants $a \in \mathbb{R}$ and $b \in \mathbb{N}$ by setting

$$s^{a,b}(x,y) = a - \|x-y\|_2^b$$

for $x, y \in \mathcal{F}$. In the following, we consider the case $a = 2$ and $b = 2$. Furthermore, assume that x and y are normalized with respect to the ℓ^2-norm. Show that, in this case, the measure $s^{a,b}$ is simply twice the inner product $\langle x|y\rangle$, which measures the cosine of the angle between x and y.

Solution to Exercise 4.1. Assuming that x and y are normalized with respect to the ℓ^2-norm, we obtain $1 = \|x\|_2^2 = \langle x|x\rangle$ and $1 = \|y\|_2^2 = \langle y|y\rangle$. From this, it follows that

$$\begin{aligned} \|x-y\|_2^2 &= \langle x-y|x-y\rangle \\ &= \langle x|x\rangle - 2\langle x|y\rangle + \langle y|y\rangle \\ &= 2 - 2\langle x|y\rangle. \end{aligned}$$

Exercise 4.2. In (4.11), we have introduced a forward smoothing procedure. This procedure results in a fading out of the paths, in particular when using a large length parameter. To avoid this fading out, one idea is to additionally apply the averaging filter in backward direction. The final self-similarity matrix is then obtained by taking the cell-wise maximum over the forward-smoothed and backward-smoothed matrices. Formalize this procedure by giving a mathematical description. Furthermore, show how the backward smoothing can be realized by forward smoothing considering the time-reversed feature sequence.
[**Hint:** To avoid boundary considerations, assume that $\mathbf{S}$ is suitably zero-padded. The effect of the forward–backward smoothing procedure is illustrated by Figure 4.12d. Another example is shown in Figure 4.15c.]

Solution to Exercise 4.2. Let L be the length parameter. As formalized by (4.11), forward smoothing is given by

$$\mathbf{S}_L^{\mathrm{For}}(n,m) := \frac{1}{L}\sum_{\ell=0}^{L-1} \mathbf{S}(n+\ell, m+\ell),$$

for $n,m \in [1:N]$ assuming that $\mathbf{S}$ is suitably zero-padded. Similarly, backward smoothing is given by

$$\mathbf{S}_L^{\mathrm{Back}}(n,m) := \frac{1}{L}\sum_{\ell=0}^{L-1} \mathbf{S}(n-\ell, m-\ell).$$

Combining forward and backward smoothing, the final self-similarity matrix is given by

$$\mathbf{S}_L^{\mathrm{Comb}}(n,m) := \max\left(\mathbf{S}_L^{\mathrm{For}}(n,m), \mathbf{S}_L^{\mathrm{Back}}(n,m)\right).$$

The matrix $\mathbf{S}_L^{\mathrm{Back}}$ can also be obtained as follows. First revert the feature sequence $X = (x_1, x_2, \ldots, x_N)$ to obtain $X^{\mathrm{Rev}} = (x_N, \ldots, x_2, x_1)$. Then compute a self-similarity matrix $\mathbf{S}^{\mathrm{Rev}}$ from X^{Rev}. Apply forward smoothing to $\mathbf{S}^{\mathrm{Rev}}$ to obtain $\mathbf{S}_L^{\mathrm{Rev,For}}$. Finally, reverting the matrix $\mathbf{S}_L^{\mathrm{Rev,For}}$ in both directions (horizontally as well as vertically) one obtains $\mathbf{S}_L^{\mathrm{Back}}$.

Exercise 4.3. Let $\mathcal{F} = \mathbb{R}^D$ as in Exercise 4.1 and $s : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ be the similarity measure defined by $s(x,y) := |\langle x|y\rangle|$ for $x,y \in \mathcal{F}$ (see (4.3)). Show that the transposition-invariant self-similarity matrix $\mathbf{S}^{\mathrm{TI}}$ (see (4.15)) is symmetric. Is the transposition index matrix $\mathbf{I}$ (see (4.16)) symmetric? Describe the relation between the matrix $\mathbf{I}$ and its transposed matrix $\mathbf{I}^\top$.

Solution to Exercise 4.3. First note that the inner product is symmetric, which implies $s(x,y) = s(y,x)$ for $x,y \in \mathcal{F}$. Furthermore, the inner product is invariant under cyclic rotations, which implies $s(\rho^i(x), \rho^i(y)) = s(x,y)$ for $i \in \mathbb{Z}$ and $x,y \in \mathcal{F}$. Now, let $X = (x_1, \ldots, x_N)$ be a feature sequence. From the above two properties and (4.14), it follows that

$$\begin{aligned}
\rho^i(\mathbf{S})(n,m) &= s(\rho^i(x_n), y_m) \\
&= s(x_n, \rho^{-i}(y_m)) \\
&= s(\rho^{-i}(y_m), x_n) \\
&= \rho^{-i}(\mathbf{S})(m,n)
\end{aligned}$$

for $i \in \mathbb{Z}$ and $n,m \in [1:N]$. In other words, a maximizing index i at coordinate (n,m) induces a maximizing index $-i$ (or $(-i \bmod 12)$) at coordinate (m,n). Using the definition in (4.15), we obtain

$$\begin{aligned}\mathbf{S}^{\mathrm{TI}}(n,m) &= \max_{i\in[0:11]} \rho^{i}(\mathbf{S})(n,m)\\ &= \max_{i\in[0:11]} \rho^{-i}(\mathbf{S})(m,n)\\ &= \max_{i\in[0:11]} \rho^{i}(\mathbf{S})(m,n)\\ &= \mathbf{S}^{\mathrm{TI}}(m,n).\end{aligned}$$

This shows that $\mathbf{S}^{\mathrm{TI}}$ is symmetric. Furthermore, we have seen that

$$\mathbf{I}^{\top}(n,m) = \mathbf{I}(m,n) = (-\mathbf{I}(n,m) \bmod 12).$$

In particular, this shows that the transposition index matrix $\mathbf{I}$ is in general not symmetric.

Exercise 4.4. For computing the matrix $\mathbf{S}_{L,\Theta}$ in (4.13), a set Θ of relative tempo differences needs to be specified. Assume that $\theta_{\min}$ is a lower bound and $\theta_{\max}$ is an upper bound for the expected relative tempo differences. For a given number $K \in \mathbb{N}$, determine a set

$$\Theta = \{\theta_1 = \theta_{\min}, \theta_2, \ldots, \theta_{K-1}, \theta_K = \theta_{\max}\}$$

consisting of increasing tempo values that are logarithmically spaced. Write a small computer program for computing this set for the parameters $\theta_{\min} = 0.66$, $\theta_{\max} = 1.5$, and $K = 5$, as well as for $\theta_{\min} = 0.5$, $\theta_{\max} = 2$, and $K = 7$.
[**Hint:** Convert the tempo bounds $\theta_{\min}$ and $\theta_{\max}$ into the log domain by applying a logarithm. Then, linearly sample the resulting interval using K samples and apply an exponential function to the samples.]

Solution to Exercise 4.4. The following program (in MATLAB) computes the tempo values for the paramters $\theta_{\min} = 0.66$, $\theta_{\max} = 1.5$, and $K = 5$:

```
tempoNum = 5; tempoMin = 0.66; tempoMax = 1.5;
logTempoMin = log10(tempoMin);
logTempoMax = log10(tempoMax);
logTempo = linspace(logTempoMin,logTempoMax,tempoNum);
tempo = 10.^logTempo;
```

From this, one obtains the following result:

$$\Theta = \{0.6600, 0.8104, 0.9950, 1.2217, 1.5000\}.$$

For the paramters $\theta_{\min} = 0.5$, $\theta_{\max} = 2$, and $K = 7$, one obtains

$$\Theta = \{0.5000, 0.6300, 0.7937, 1.0000, 1.2599, 1.5874, 2.0000\}.$$

Exercise 4.5. In this exercise, we look at the various thresholding strategies introduced in Section 4.2.2.4. Given the matrix

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 4 & 3 & 4 & 3 \\ 1 & 1 & 2 & 2 \\ 5 & 6 & 6 & 5 \end{bmatrix},$$

compute the matrices that are obtained by applying the following thresholding operations:

(a) Global thresholding using $\tau = 4$
(b) Global thresholding using $\tau = 4$ as in (a) with subsequent linear scaling of the range $[\tau, \mu]$ to $[0,1]$ using $\mu := \max\{\mathbf{S}(n,m) \mid n,m \in [1:4]\}$
(c) Global thresholding with subsequent linear scaling as in (b) and applying the penalty parameter $\delta = -1$
(d) Relative thresholding using the relative threshold parameter $\rho = 0.5$
(e) Local thresholding in a column- and rowwise fashion using $\rho = 0.5$

Solution to Exercise 4.5.

(a)
$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 4 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \\ 5 & 6 & 6 & 5 \end{bmatrix}$$

(b)
$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0.5 & 1 & 1 & 0.5 \end{bmatrix}$$

(c)
$$\begin{bmatrix} -1 & -1 & -1 & -1 \\ 0 & -1 & 0 & -1 \\ -1 & -1 & -1 & -1 \\ 0.5 & 1 & 1 & 0.5 \end{bmatrix}$$

(d)
$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 4 & 3 & 4 & 3 \\ 0 & 0 & 0 & 0 \\ 5 & 6 & 6 & 5 \end{bmatrix}$$

(e)
$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 4 & 0 & 4 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 6 & 6 & 0 \end{bmatrix}$$

Exercise 4.6. Let $X = (x_1, x_2, \ldots, x_N)$ be a sequence and $\alpha = [s:t] \subseteq [1:N]$ a segment of length $M := |\alpha|$. Show that the optimization procedure for computing an optimal path family over α (as described in Section 4.3.1.2) has a complexity of $O(MN)$ regarding the memory requirements as well as the running time.

Solution to Exercise 4.6. In the procedure, one needs to compute the submatrix $\mathbf{S}^{\alpha} \in \mathbb{R}^{N,M}$, the accumulated score matrix $\mathbf{D} \in \mathbb{R}^{N,M+1}$, and the optimal path family $\mathcal{P}^*$. Obviously, all these objects can be stored using $O(MN)$ real numbers. The most expensive part for computing $\mathbf{D}$ is the recursion (4.26). Note that the set of predecessors $\Phi(n,m)$ contains at most three elements for each cell (n,m). Therefore, to compute the recursion (4.26), one requires one addition and the maximization over a set that contains at most three elements for each cell (n,m), $n \in [2:N]$, $m \in [2:M]$. Altogether this requires $O(MN)$ operations. The compuation of the remaining values of $\mathbf{D}$ (initialization) obviously requires less operations. Finally, the backtracking procedure for computing $\mathcal{P}^*$ requires a number of operations that is linear in the total length of $\mathcal{P}^*$—a number that depends linearly on N and M. This proves the claim.

Exercise 4.7. Let $X = (x_1, \ldots, x_N)$ be a feature sequence and $\mathbf{S}$ the resulting SSM satisfying the normalization properties (4.18) and (4.19). Let $\mathcal{P}^*$ be an optimal path family over a given segment α. Show that $|\alpha| \leq \sigma(\mathcal{P}^*) \leq N$. In particular, this shows that $\sigma(\mathcal{P}^*) = N$ for $\alpha = [1:N]$.

Solution to Exercise 4.7. For $\alpha = [s:t]$, let $P_0 = \big((s,s),(s+1,s+1),\ldots,(t,t)\big)$ be the path over α running along the main diagonal. Let $\mathcal{P}_0 := \{P_0\}$ be the path familiy over α consisting of the single path P_0. The normalization property (4.19) implies that $\sigma(\mathcal{P}_0) = \sigma(P_0) = |\alpha|$. This shows that $\sigma(\mathcal{P}^*) \geq \sigma(\mathcal{P}_0) = |\alpha|$ for an optimal path family $\mathcal{P}^*$ (see (4.24)).

Because of the step size condition $\Sigma = \{(2,1),(1,2),(1,1)\}$ used in the procedure (see Section 4.3.1.2), it follows that the total length of any path family is at most N. From the normalization property (4.18), it follows that $\sigma(\mathcal{P}^*) \leq N$.

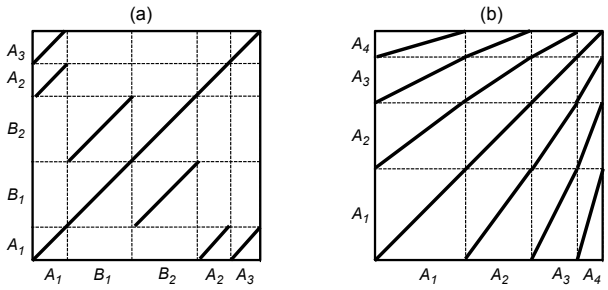
Exercise 4.8. For two given real numbers $a,b \in \mathbb{R}$, the arithmetic mean is defined by $A(a,b) = (a+b)/2$, the geometric mean by $G(a,b) = \sqrt{ab}$, and the harmonic mean by $H(a,b) = 2ab/(a+b)$. Show that $H(a,b) \leq G(a,b) \leq A(a,b)$, i.e., the geometric mean always lies between the harmonic mean and the arithmetic mean. Furthermore, compute $A(a,b)$, $G(a,b)$, and $H(a,b)$ for the numbers $a = 1$ and $b \in \{1,2,3,4\}$.

Solution to Exercise 4.8. For $a,b \in \mathbb{R}$, we have $(a-b)^2 \geq 0$, which implies $a^2 + 2ab + b^2 \geq 4ab$, and hence $(a+b)^2/4 \geq ab$. Taking the square root on both sides, this implies $A(a,b) \geq G(a,b)$. Furthermore, $(a+b)^2 \geq 4ab$ yields $ab \geq 4a^2b^2/(a+b)^2$. Again, taking the square root on both sides, this implies $G(a,b) \geq H(a,b)$. For the numbers $a = 1$ and $b \in \{1,2,3,4\}$, one obtains the following values for $A(a,b)$, $G(a,b)$, and $H(a,b)$:

a	b	$A(a,b)$	$G(a,b)$	$H(a,b)$
1	1	1	1	1
1	2	1.5	$\sqrt{2} \approx 1.41$	$4/3 \approx 1.33$
1	3	2	$\sqrt{3} \approx 1.73$	1.5
1	4	2.5	2	1.6

Exercise 4.9. (a) Let us consider a piece of music having the musical structure $A_1B_1B_2A_2A_3$, where we assume that corresponding parts are repeated in exactly the same way. Furthermore, assume that the A-part and B-part segments are completely unrelated to each other and that a B-part segment has exactly twice the length of an A-part segment. Sketch an idealized SSM for this piece (as in Figure 4.18). Furthermore, determine the fitness values of the segments corresponding to A_1 and B_1, respectively.
(b) Next, consider a piece having the musical structure $A_1A_2A_3A_4$, where the four parts are repeated with increasing tempo. Assume that A_1 lasts 20 seconds, A_2 lasts 15 seconds, A_3 lasts 10 seconds, and A_4 lasts 5 seconds. Again sketch an idealized SSM and determine the fitness values of the four segments corresponding to the four parts.

Solution to Exercise 4.9. The following figure shows the idealized SSMs of the cases (a) and (b):



We assume that the idealized SSMs have the value one on the indicated paths and otherwise the value zero (similar to Figure 4.18). Then, for the segments in question, the normalized score coincides with the normalized coverage, thus also coinciding with the fitness value. As for **(a)**, one obtains $\gamma(\alpha) = 3/7$, $\bar{\gamma}(\alpha) = 2/7$, and $\varphi(\alpha) = 2/7$ for α corresponding to A_1. Furthermore, one obtains $\gamma(\alpha) = 4/7$, $\bar{\gamma}(\alpha) = 2/7$, and $\varphi(\alpha) = 2/7$ for α corresponding to B_1.

As for **(b)**, one obtains $\varphi(\alpha) = 3/5$ for α corresponding to A_1, $\varphi(\alpha) = 7/10$ for α corresponding to A_2, $\varphi(\alpha) = 4/5$ for α corresponding to A_3, and $\varphi(\alpha) = 9/10$ for α corresponding to A_4.

Exercise 4.10. Let $[1:N]$ be a sampled time axis. Show that the number of different segments $\alpha = [s:t]$ with $s,t \in [1:N]$ and $s \leq t$ is $(N+1)N/2$.

Solution to Exercise 4.10. There is one segment of length N, two segments of length $N-1$, three segments of length $N-2$, and so on, and finally N segments of length 1. Therefore, altogether, the number of segments is $\sum_{n=1}^{N} n = (N+1)N/2$.

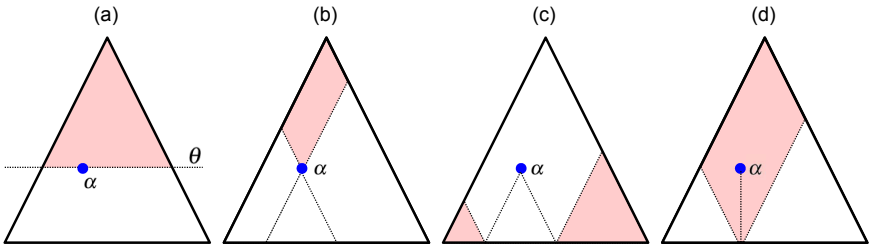
Exercise 4.11. Determine the overall computational complexity of calculating the fitness scape plot as introduced in Section 4.3.2 for a feature sequence $X = (x_1, x_2, \ldots, x_N)$ of length N.
[**Hint:** Use Exercise 4.6 and Exercise 4.10.]

Solution to Exercise 4.11. In Exercise 4.6, we showed that the computational complexity for calculating the fitness of a segment α of length $M = |\alpha|$ is $O(MN)$. To derive the scape plot, one needs to calculate the fitness of all possible segments. By Exercise 4.10, there are $(N+1)N/2$ segments. Since $M \leq N$, this shows that the overall computation complextity is $O(N^4)$.

Exercise 4.12. Given a triangular representation of all segments within $[1:N]$ as in Figure 4.19b, visually indicate the following sets of segments:

(a) All segments having a minimal length above a given threshold $\theta \geq 0$
(b) All segments that contain a given segment α
(c) All segments that are disjoint to a given segment α
(d) All segments that contain the center $c(\alpha)$ of a given segment α

Solution to Exercise 4.12. In the following triangular representations, the relevant sets of segments are indicated by the colored regions:

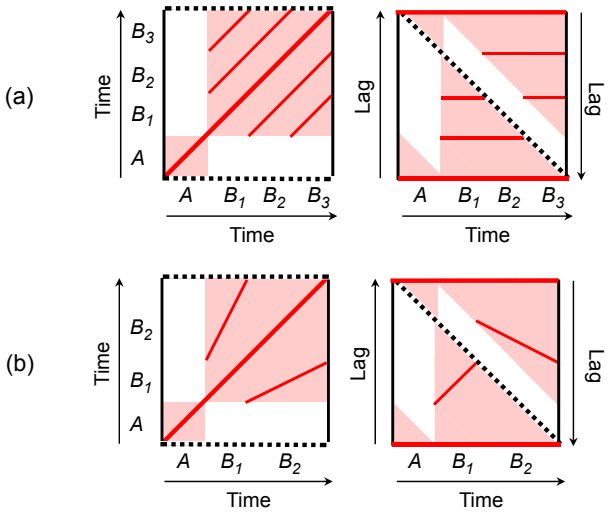


Exercise 4.13. Sketch the similarity matrix **S** and the circular time-lag matrix $\mathbf{L}^{\circ}$ as in Figure 4.26c for pieces with the following musical structure:

(a) $AB_1B_2B_3$, where all segments have the same length

(b) AB_1B_2, where the A-part and B_1-part segments have the same length and the B_2-part segment has twice the length (played with half the tempo of B_1)

Solution to Exercise 4.13. The following figure shows the similarity matrix **S** (left) and the circular time-lag matrix $\mathbf{L}^{\circ}$ (right) for the two cases:



Exercise 4.14. Sketch the similarity matrix **S**, the circular time-lag matrix $\mathbf{L}^{\circ}$, and the resulting novelty function $\Delta_{\text{Structure}}$ for pieces with the following musical structure (assuming that all segments corresponding to a musical part have the same length and that the kernel size used for computing the novelty function is much smaller than this length):

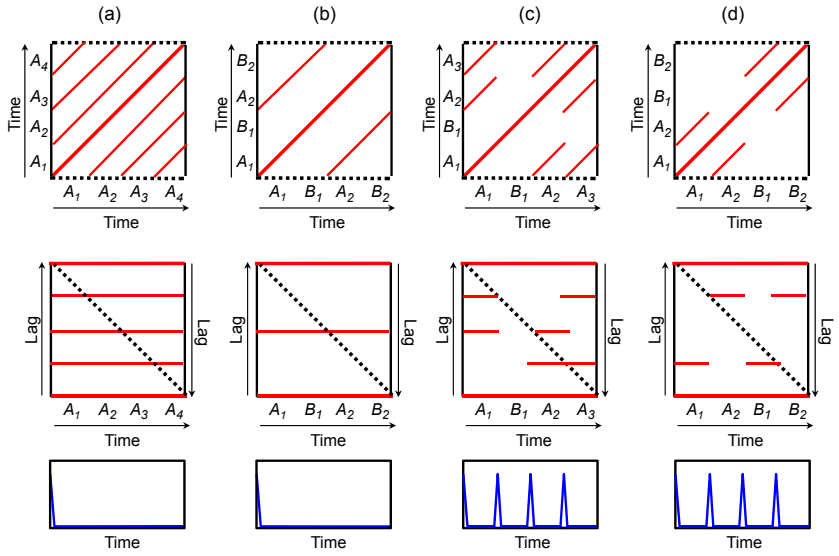
(a) $A_1A_2A_3A_4$

(b) $A_1B_1A_2B_2$

(c) $A_1B_1A_2A_3$

(d) $A_1A_2B_1B_2$

Solution to Exercise 4.14. The following figure shows the similarity matrix $\mathbf{S}$ (top), the circular time-lag matrix $\mathbf{L}^{\circ}$ (middle), and the novelty function $\Delta_{\text{Structure}}$ (bottom) for the four cases:



Exercise 4.15. Let $\mathcal{A} = \{\alpha_1, \alpha_2, \ldots, \alpha_K\}$ be a segment family together with a labeling $\lambda_k \in \Lambda$, $k \in [1:K]$. Let $\mu(\mathcal{A}) := \bigcup_{k=1}^{K} \alpha_k$ be the union of all segments. Show that one may assume $\mu(\mathcal{A}) = [1:N]$ by suitably extending the segment family, the label set Λ, and the labeling.

Solution to Exercise 4.15. Regard each frame $n \in [1:N] \setminus \mu(\mathcal{A})$ as an additional segment of length 1, which keeps the disjointness condition of a segment family. Furthermore, extend the label set by introducing for each such frame n an additional label $\lambda_n \notin \Lambda$. Assign to the additional segment corresponding to this frame the label λ_n. As an alternative, one may use only one additional label with the meaning "unannotated" and use this label for all frames $\lambda_n \notin \Lambda$. However, depending on the evaluation measure, the two extensions (one with individual labels and one with a joint label) may lead to different results.

Exercise 4.16. In (4.51), we defined the set $\mathcal{I} = \{(n,m) \in [1:N] \times [1:N] \mid n < m\}$ to serve as a set of items for defining the pairwise evaluation measure. Determine the size of $\mathcal{I}$. Furthermore, let $\varphi : [1:N] \to \Lambda$ be a label function, and let $\mathcal{I}_{+}^{\text{Ref}} = \{(n,m) \in \mathcal{I} \mid \varphi(n) = \varphi(m)\}$ be the set of positive items with regard to φ. Derive a general formula for the size of $\mathcal{I}_{+}^{\text{Ref}}$.
[**Hint:** Note that the size of $\mathcal{I}_{+}^{\text{Ref}}$ does not depend on the original order of the frames. Given a specific label, consider the number of frames assigned to that label. To derive a formula for the size of $\mathcal{I}_{+}^{\text{Ref}}$, one needs to consider all possible labels assumed by φ.]

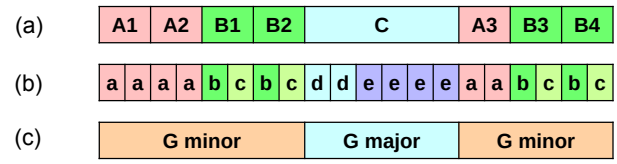
Solution to Exercise 4.16. The size of $\mathcal{I}$ corresponds to the number of two-element subsets of $[1:N]$. This number is given by

$$|\mathcal{I}| = \binom{N}{2} = \frac{N(N-1)}{2}.$$

Suppose that φ assumes K different values, and let $\lambda_1, \ldots, \lambda_K \in \Lambda$ be these values. Furthermore, let $N_k := |\varphi^{-1}(\lambda_k)|$ be the number of frames assigned to the label λ_k, $k \in [1:K]$. Then there are $\binom{N_k}{2}$ pairs $(n,m) \in \mathcal{I}$ with $\varphi(n) = \varphi(m) = \lambda_k$. Therefore, the size of $\mathcal{I}_+^{\mathrm{Ref}}$ is given by

$$|\mathcal{I}_+^{\mathrm{Ref}}| = \sum_{k\in[1:K]} \binom{N_k}{2} = \sum_{k\in[1:K]} \frac{N_k(N_k-1)}{2}.$$

Exercise 4.17. In this exercise, we investigate how the pairwise labeling evaluation behaves with respect to under- and oversegmentation. To this end, let us consider the following structure annotations of a piece of music (similar to our Brahms example shown in Figure 4.28):



Compute the size $|\mathcal{I}_+|$ for each of the three annotations. Then, assume that (a) is the reference annotation. Compute the pairwise precision, recall, and F-measure for the case that (b) is the estimated annotation ("oversegmentation") and for the case that (c) is the estimated annotation ("undersegmentation").
[**Hint:** Use the results of Exercise 4.16.]

Solution to Exercise 4.17. In this example, we have $N = 20$. By Exercise 4.16, the number of positives with respect to annotation (a), (b), and (c) are:

$$\text{(a)}: \quad |\mathcal{I}_+^{(\mathrm{a})}| = \binom{6}{2} + \binom{8}{2} + \binom{6}{2} = 15 + 28 + 15 = 58,$$

$$\text{(b)}: \quad |\mathcal{I}_+^{(\mathrm{b})}| = \binom{6}{2} + \binom{4}{2} + \binom{4}{2} + \binom{2}{2} + \binom{4}{2} = 15 + 6 + 6 + 1 + 6 = 34,$$

$$\text{(c)}: \quad |\mathcal{I}_+^{(\mathrm{c})}| = \binom{14}{2} + \binom{6}{2} = 91 + 15 = 106.$$

First, let us compare annotation (b) against the reference annotation (a). Since (b) is a refinement of (a), all positive items with regard to (b), are also positive with regard to (a). Therefore, #TP = 34, #FP = 0, and #FN = 24. From this, one obtains P = 1,

$R = 34/58 \approx 0.586$, and $F \approx 0.739$. Next, let us compare annotation (c) against the reference annotation (a). Since (c) is a coarsening of (a), all positive items with regard to (a) are also positive with regard to (c). Therefore, #TP $= 58$, #FP $= 48$, and #FN $= 0$. From this, one obtains $P = 58/106 \approx 0.547$, $R = 1$, and $F \approx 0.707$.

Exercise 4.18. Let $[1:N]$ be a sampled time axis with $N = 50$. Furthermore, let $B^{\mathrm{Ref}} = \{7,13,19,28,40,44\}$ be a reference boundary annotation and $B^{\mathrm{Est}} = \{6,12,21,29,42\}$ be an estimated boundary annotation. Compute the boundary evaluation measures (precision, recall, F-measure) as in Section 4.5.4 for the tolerance parameter $\tau = 0$, $\tau = 1$, and $\tau = 2$, respectively. Why is the case $\tau = 2$ problematic for this example?

Solution to Exercise 4.18. In the case $\tau = 0$, none of the boundaries agree resulting in $P = R = F = 0$. In the case $\tau = 1$, one has #TP $= 3$, #FP $= 2$, and #FN $= 3$, which yields $P = 3/5$, $R = 1/2$, and $F = 6/11$. In the case $\tau = 2$, one obtains $P = R = F = 1$. This case is problematic, since the boundary $b = 42$ of the estimated annotation agrees with the two boundaries $b = 40$ and $b = 44$ of the reference annotation. Note that the condition $|b_{k+1} - b_k| > 2\tau$ of (4.58) is violated for the reference annotation, where one one obtains $|b_{k+1} - b_k| = |44 - 40| = 4$ for $k = 5$ and $2\tau = 4$ for $\tau = 2$.

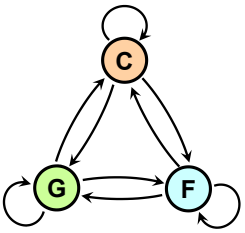
Exercise 4.19. Let $[1:N]$ be a sampled time axis with $N = 100$. Furthermore, let $\mathcal{A}^{\mathrm{Ref}} = \{[16:26],[40:49],[50:60],[75:84]\}$ be a reference thumbnail family. Compute the thumbnail F-measure as introduced in Section 4.5.5 for the following estimated thumbnail segments:

(a) $\alpha^{\mathrm{Est}} = [18:27]$
(b) $\alpha^{\mathrm{Est}} = [45:54]$
(c) $\alpha^{\mathrm{Est}} = [60:75]$

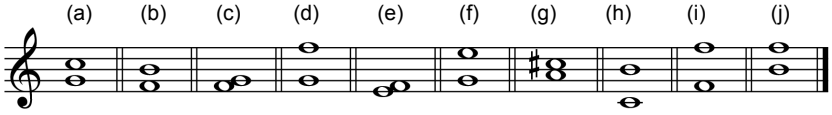
Solution to Exercise 4.19. First note that the F-measure between two nonoverlapping segments is zero.

(a) The only segment of $\mathcal{A}^{\mathrm{Ref}}$ overlapping with estimated thumbnail segment $\alpha^{\mathrm{Est}} = [18:27]$ is $\alpha = [16:26]$. For these two segments, one obtains $P^{\alpha} = 9/10$, $R^{\alpha} = 9/11$, and $F^{\mathrm{Thumb}} = F^{\alpha} = 6/7 \approx 0.857$.
(b) The estimated thumbnail segment $\alpha^{\mathrm{Est}} = [45:54]$ overlaps with the two segments $[40:49]$ and $[50:60]$ of $\mathcal{A}^{\mathrm{Ref}}$. For $\alpha = [40:49]$, one obtains $P^{\alpha} = 5/10$, $R^{\alpha} = 5/10$, and $F^{\alpha} = 1/2$. For $\alpha = [50:60]$, one obtains $P^{\alpha} = 5/10$, $R^{\alpha} = 5/11$, and $F^{\alpha} = 10/21$. Therefore, one obtains $F^{\mathrm{Thumb}} = 1/2$.
(c) The estimated thumbnail segment $\alpha^{\mathrm{Est}} = [60:75]$ overlaps with the two segments $[50:60]$ and $[75:84]$ in a single frame, respectively. The F-measure will be higher for the shorter segment $\alpha = [75:84]$, from which obtains $P^{\alpha} = 1/16$, $R^{\alpha} = 1/10$, and $F^{\mathrm{Thumb}} = F^{\alpha} = 1/13$.

Chapter 5
Chord Recognition



Exercise 5.1. Determine, for each of the following intervals, the number of semitones and the interval name (as specified in Figure 5.3):

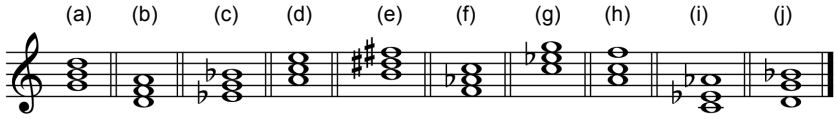


Solution to Exercise 5.1. The number of semitones is given in brackets []. **(a)** Perfect fourth [5]. **(b)** Tritone [6]. **(c)** Major second [2]. **(d)** Minor seventh [10]. **(e)** Minor second [1]. **(f)** Major sixth [9]. **(g)** Major third [4]. **(h)** Major seventh [11]. **(i)** Perfect octave [12]. **(j)** Tritone [6].

Exercise 5.2. The **complement** of an interval is the interval which, when added to the original interval, spans an octave in total. Specify the complement for each interval in Figure 5.3. In which way is the tritone interval special?

Solution to Exercise 5.2. One obtains the following pairs of complementary intervals: perfect unison – perfect octave, minor second – major seventh, major second – minor seventh, minor third – major sixth, major third – minor sixth, perfect fourth – perfect fifth, tritone – tritone. The tritone interval is special in the way that it is the only interval that coincides with its complement. It splits the octave in exactly two halves.

Exercise 5.3. Determine the chord symbol for each of the following chords (similar to Figure 5.6):



Solution to Exercise 5.3. (a) G (b) Dm (c) E$^\flat$ (d) Am (e) B (f) Fm (g) Cm (h) F (i) A$^\flat$ (j) Gm

Exercise 5.4. In this exercise, we compare the size of the intervals obtained from different definitions. First, assuming the twelve-tone equal-tempered scale, determine the size (given in cents) and frequency ratios for each of the 13 intervals shown in Figure 5.3. Next, assuming just intonation, determine the size (given in

cents) and the frequency ratios of the intervals (see Figure 5.3). Finally, compute the difference of the interval sizes (given in cents) between the just intonation and the equal-tempered case.
[**Hint:** Write a small computer program that helps you with the calculations.]

Solution to Exercise 5.4. In the twelve-tone equal-tempered scale, a semitone corresponds to 100 cents. Therefore, an interval consisting of Δ semitones has the size $\Delta \cdot 100$ cents (see Section 1.3.2). Furthermore, the frequency ratio of an interval consisting of Δ semitones is given by $1 : 2^{\Delta/12}$.

The frequency ratios of the intervals with respect to just intonation are specified in Figure 5.3. Applying (1.4), one obtains the interval sizes given in cents. The following table yields the results:

Interval name	Interval	Equal-tempered scale		Just intonation		Difference
		Size (cents)	Freq. ratio	Size (cents)	Freq. ratio	(cents)
(Perfect) unison	C4–C4	0	1 : 1.0000	0.0	1 : 1	0.0000
Minor second	C4–D♭4	100	1 : 1.0595	111.7	15 : 16	11.7313
Major second	C4–D4	200	1 : 1.1225	203.9	8 : 9	3.9100
Minor third	C4–E♭4	300	1 : 1.1892	315.6	5 : 6	15.6413
Major third	C4–E4	400	1 : 1.2599	386.3	4 : 5	-13.6863
(Perfect) fourth	C4–F4	500	1 : 1.3348	498.0	3 : 4	-1.9550
Tritone	C4–F♯4	600	1 : 1.4142	590.2	32 : 45	-9.7763
(Perfect) fifth	C4–G4	700	1 : 1.4983	702.0	2 : 3	1.9550
Minor sixth	C4–A♭4	800	1 : 1.5874	813.7	5 : 8	13.6863
Major sixth	C4–A4	900	1 : 1.6818	884.4	3 : 5	-15.6413
Minor seventh	C4–B♭4	1000	1 : 1.7818	1017.6	5 : 9	17.5963
Major seventh	C4–B4	1100	1 : 1.8877	1088.3	8 : 15	-11.7313
(Perfect) octave	C4–C5	1200	1 : 2.0000	1200.0	1 : 2	0.0000

Exercise 5.5. In this exercise, we investigate the dependency between the degree of consonance of an interval and the coincidence of partials of the two notes underyling the interval. Assuming the twelve-tone equal-tempered scale, we look at the intervals that are formed by the root note C4 and each of the following seven notes: C4, E♭4, E4, F4, F♯4, G4, and C5. Consider for each of these notes the first eight harmonics. Determine for each of the resulting harmonics the closest musical note along with the difference (given in cents) between the harmonic's actual frequency and the center frequency of the musical note (see also Figure 1.20). For example, the following table shows these results for the two notes C4 and E♭4 (with the differences being specified in brackets):

1	2	3	4	5	6	7	8
C4 [0]	C5 [0]	G5 [+2]	C6 [0]	E6 [-14]	G6 [+2]	B♭6 [-31]	C7 [0]
E♭4 [0]	E♭5 [0]	B♭5 [+2]	E♭6 [0]	G6 [-14]	B♭6 [+2]	D♭6 [-31]	E♭7 [0]
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

Then investigate, for each of the seven intervals, which of the harmonics of the two involved notes coincide (or, to be more precise, are close together with respect to

frequency). For example, the coincidences of harmonics between the notes of the interval C4–C4 and the interval C4–E♭4 are indicated by frameboxes in the above table (where the first row represents the interval C4–C4 and the second one the interval C4–E♭4). Note that G6 appears as the sixth harmonic of C4 and as the fifth harmonic of E♭4. However, this coincidence is "tarnished" by the fact that the sixth harmonic of C4 deviates by $+2$ cents from the center frequency of G6, whereas the fifth harmonic of E♭4 deviates by -14 cents from G6. Similarly, discuss the results for the other intervals.

Solution to Exercise 5.5. The following table specifies the first eight harmonics for each of the seven notes (listed in the first column, corresponding to the first harmonics). The differences between the harmonics' frequencies and the respective notes' center frequencies are given by Figure 1.20. The values (given in cents) are specified in brackets following the respective musical notes.

1	2	3	4	5	6	7	8
C4 [0]	C5 [0]	G5 [+2]	C6 [0]	E6 [-14]	G6 [+2]	B♭6 [-31]	C7 [0]
E♭4 [0]	E♭5 [0]	B♭5 [+2]	E♭6 [0]	G6 [-14]	B♭6 [+2]	D♭6 [-31]	E♭7 [0]
E4 [0]	E5 [0]	B5 [+2]	E6 [0]	G♯6 [-14]	B6 [+2]	D6 [-31]	E7 [0]
F4 [0]	F5 [0]	C5 [+2]	F6 [0]	A6 [-14]	C6 [+2]	E♭6 [-31]	F7 [0]
F♯4 [0]	F♯5 [0]	C♯5 [+2]	F♯6 [0]	A♯6 [-14]	C♯6 [+2]	E6 [-31]	F♯7 [0]
G4 [0]	G5 [0]	D5 [+2]	G6 [0]	B6 [-14]	D7 [+2]	F7 [-31]	G7 [0]
C5 [0]	C6 [0]	G6 [+2]	C7 [0]	E7 [-14]	G7 [+2]	B♭7 [-31]	C8 [0]

The coincidence of harmonics between the notes of the seven intervals are indicated by frameboxes. Obviously, the (perfect) unison C4–C4 is the most consonant interval, where all eight harmonics of both notes agree. Next, the (perfect) octave C4–C5 is highly consonant, where the harmonics of C5 are also harmonics of C4. The (perfect) fifth C4–G4 is consonant. For example, the note G5 appears as the third harmonic of C4 and as the second harmonic of G4. Similarly, G6 appears as the sixth harmonic of C4 and as the fourth harmonic of G4.

As a typical dissonant interval, let us consider the tritone C4–F♯4. For this interval, there is hardly any overlap in the harmonics. The only note appearing in the first eight harmonics of C4 and F♯4 is the note E6, which appears as the fifth harmonic of C4 and as the seventh harmonic of F♯4. However, this relation is "tarnished" by the fact that the fifth harmonic of C4 deviates by -14 cents from E6, whereas the seventh harmonic of F♯4 deviates by -31 from E6. Therefore, despite coinciding on the note level, these two harmonics do not really agree on the frequency level.

Exercise 5.6. In Figure 5.20b, one can observe many misclassifications and chord label changes in the recognition result. Explain why these errors only occur in the second and third measure, while the first and fourth measure have been classified correctly.

Solution to Exercise 5.6. As explained in Section 5.2.3.4, the misclassifications and chord label changes are the result of using short analyis frames (200 ms), which are dominated by the sound of only one or two notes. Now, the first and fourth measures are labeled as **C**. Also, the first note of each broken chord in these measures is a C, the root of **C**. These notes are held throughout the duration of the respective broken chord. Recall from Section 5.2.3.2 that the harmonics of the single note C already produce a chroma pattern close to **C**. This yields an explanation of why the broken chords in the first and fourth measure have been correctly labeled as **C**, even for frames where not all notes of the chord have been active. As for the second and third measures, the situation is different. In these measures, the first notes do not coincide with the root notes of the corresponding chords. For example, in the second measure, the first note is a C, whereas the chord is **Dm**. In the third measure, the first note is a B, whereas the chord is **G**. This leads to local misclassifications.

Exercise 5.7. Let Λ be the set of the major and minor triads (see (5.5)). Furthermore, for a given chord $\lambda \in \Lambda$, let $\mathbf{t}^{\mathrm{h}}_{\lambda}$ be the chord template with harmonics based on the first eight harmonics (see (5.13) and (5.14)). Compute $\mathbf{t}^{\mathrm{h}}_{\lambda}$ for $\lambda = \mathbf{C}$ and $\lambda = \mathbf{Cm}$, respectively, using the parameter $\alpha = 1$.

Solution to Exercise 5.7. The major triad $\lambda = \mathbf{C}$ consists of three notes belonging to the pitch classes C, E, and G. For each of these pitch classes, the following table specifies the template with harmonics according to (5.13) using the parameter $\alpha = 1$ as well as the chord template $\mathbf{t}^{\mathrm{h}}_{\mathbf{C}} = \mathbf{t}^{\mathrm{h}}_{\mathrm{C}} + \mathbf{t}^{\mathrm{h}}_{\mathrm{E}} + \mathbf{t}^{\mathrm{h}}_{\mathrm{G}}$ according to (5.14):

	C	C$^\sharp$	D	D$^\sharp$	E	F	F$^\sharp$	G	G$^\sharp$	A	A$^\sharp$	B
$\mathbf{t}^{\mathrm{h}}_{\mathrm{C}}$	4	0	0	0	1	0	0	2	0	0	1	0
$\mathbf{t}^{\mathrm{h}}_{\mathrm{E}}$	0	0	1	0	4	0	0	0	1	0	0	2
$\mathbf{t}^{\mathrm{h}}_{\mathrm{G}}$	0	0	2	0	0	1	0	4	0	0	0	1
$\mathbf{t}^{\mathrm{h}}_{\mathbf{C}}$	4	0	3	0	5	1	0	6	1	0	1	3

The minor triad $\lambda = \mathbf{Cm}$ consists of three notes belonging to the pitch classes C, E$^\flat$, and G (where we identify E$^\flat$ with D$^\sharp$ due to enharmonic equivalence). Similar to the major case, the following table shows the result for the chord template $\mathbf{t}^{\mathrm{h}}_{\mathbf{Cm}} = \mathbf{t}^{\mathrm{h}}_{\mathrm{C}} + \mathbf{t}^{\mathrm{h}}_{\mathrm{E}^\flat} + \mathbf{t}^{\mathrm{h}}_{\mathrm{G}}$:

	C	C$^\sharp$	D	D$^\sharp$	E	F	F$^\sharp$	G	G$^\sharp$	A	A$^\sharp$	B
$\mathbf{t}^{\mathrm{h}}_{\mathrm{C}}$	4	0	0	0	1	0	0	2	0	0	1	0
$\mathbf{t}^{\mathrm{h}}_{\mathrm{E}^\flat}$	0	1	0	4	0	0	0	1	0	0	2	0
$\mathbf{t}^{\mathrm{h}}_{\mathrm{G}}$	0	0	2	0	0	1	0	4	0	0	0	1
$\mathbf{t}^{\mathrm{h}}_{\mathbf{Cm}}$	4	1	2	4	1	1	0	7	0	0	3	1

Exercise 5.8. In this exercise, we extend the chord template model as defined by (5.13) and (5.14) by introducing some additional weight parameters. For the C-major chord $\lambda = \mathbf{C}$, we define the template

$$\mathbf{t}^{\mathrm{h,w}}_{\mathbf{C}} = w_1 \cdot \mathbf{t}^{\mathrm{h}}_{\mathrm{C}} + w_2 \cdot \mathbf{t}^{\mathrm{h}}_{\mathrm{E}} + w_3 \cdot \mathbf{t}^{\mathrm{h}}_{\mathrm{G}}$$

for $w = (w_1, w_2, w_3)^\top \in \mathbb{R}^3$. Similarly, using the same weights, we define the chord templates $\mathbf{t}_\lambda^{\mathrm{h,w}}$ for the other major and minor chords $\lambda \in \Lambda$ (see (5.5)). We now compare these new chord templates with the original binary templates $\mathbf{t}_\lambda$ (see (5.7)) using the similarity measure s as defined in (5.8). Write a small computer program to compute the similarity values $s(\mathbf{t}_{\mathbf{C}}^{\mathrm{h,w}}, \mathbf{t}_\lambda)$ and $s(\mathbf{t}_{\mathbf{Cm}}^{\mathrm{h,w}}, \mathbf{t}_\lambda)$ for all 24 major and minor chords $\lambda \in \Lambda$ using the following parameters:

(a) $\alpha = 0$ and $w = (1,1,1)$
(b) $\alpha = 1$ and $w = (1,1,1)$
(c) $\alpha = 0$ and $w = (1,0.2,1)$
(d) $\alpha = 1$ and $w = (1,0.2,1)$

In which case is there a confusion between the C-major and C-minor chord? Explain the reason for this confusion in words.

Solution to Exercise 5.8. The following table shows the distances $s(\mathbf{t}_{\mathbf{C}}^{\mathrm{h,w}}, \mathbf{t}_\lambda)$ as well as $s(\mathbf{t}_{\mathbf{Cm}}^{\mathrm{h,w}}, \mathbf{t}_\lambda)$ for $\lambda \in \Lambda$ based on the four different settings (a) to (d) used for the computation of $\mathbf{t}_\lambda^{\mathrm{h,w}}$:

	(a)		(b)		(c)		(d)	
	C	**Cm**	**C**	**Cm**	**C**	**Cm**	**C**	**Cm**
Bm	0	0	0.3499	0.1750	0	0	0.2596	0.2164
A♯m	0	0	0.1166	0.2916	0	0	0.1442	0.1875
Am	0.6667	0.3333	0.5249	0.2916	0.4851	0.4042	0.4183	0.3606
G♯m	0	0.3333	0.2333	0.2916	0	0.0808	0.1154	0.1298
Gm	0.3333	0.3333	0.5832	0.6999	0.4042	0.4042	0.6635	0.6924
F♯m	0	0	0	0.0583	0	0	0	0.0144
Fm	0.3333	0.3333	0.3499	0.2916	0.4042	0.4042	0.3750	0.3606
Em	0.6667	0.3333	0.8165	0.5249	0.4851	0.4042	0.6635	0.5914
D♯m	0	0.3333	0.0583	0.4082	0	0.0808	0.0721	0.1587
Dm	0	0	0.2333	0.1750	0	0	0.2308	0.2164
C♯m	0.3333	0	0.3499	0.1166	0.0808	0	0.1442	0.0865
Cm	0.6667	1.0000	0.5832	0.8748	0.8085	0.8893	0.7212	**0.7934**
B	0	0.3333	0.1750	0.2916	0	0.0808	0.1010	0.1298
A♯	0	0	0.2916	0.3499	0	0	0.3029	0.3173
A	0.3333	0	0.2916	0.1166	0.0808	0	0.1298	0.0865
G♯	0.3333	0.6667	0.2916	0.4666	0.4042	0.4851	0.3029	0.3462
G	0.3333	0.3333	0.6999	0.5832	0.4042	0.4042	0.6924	0.6635
F♯	0	0	0.0583	0.2333	0	0	0.0721	0.1154
F	0.3333	0.3333	0.2916	0.2916	0.4042	0.4042	0.3606	0.3606
E	0.3333	0	0.5249	0.1166	0.0808	0	0.2452	0.1442
D♯	0.3333	0.6667	0.4082	0.8165	0.4042	0.4851	0.5049	0.6058
D	0	0	0.1750	0.1166	0	0	0.1587	0.1442
C♯	0	0	0.1166	0.1166	0	0	0.0865	0.0865
C	1.0000	0.6667	0.8748	0.6999	0.8893	0.8085	0.8511	**0.8078**

The only major-minor confusion occurs in the case (d) for the chroma pattern of $\mathbf{t}_{\mathbf{Cm}}^{\mathrm{h,w}}$. In this case, the similarity value $s(\mathbf{t}_{\mathbf{Cm}}^{\mathrm{h,w}}, \mathbf{t}_{\mathbf{Cm}}) = 0.7934$ is smaller than the value $s(\mathbf{t}_{\mathbf{Cm}}^{\mathrm{h,w}}, \mathbf{t}_{\mathbf{C}}) = 0.8078$. The reason is that the minor third E♭ has been weighted by the relatively small factor of $w_2 = 0.2$, which leads to a small value in the chroma band E♭ in $\mathbf{t}_{\mathbf{Cm}}^{\mathrm{h,w}}$. On the other side, the chroma band E in $\mathbf{t}_{\mathbf{Cm}}^{\mathrm{h,w}}$ has a relatively large

value because E appears as a harmonic partial of the tonic C, which is weighted by $w_1 = 1$. Altogether, this explains the confusion.

Exercise 5.9. Let us consider a Markov chain with I states $\{\alpha_1, \alpha_2, \ldots, \alpha_I\}$ and transition probability coefficients a_{ij}, $i, j \in [1:I]$ (see (5.20)). The goal of this excercise is to determine how long the resulting system stays (on average) in a given state. To this end, consider an observation sequence $S = (\alpha_i, \ldots, \alpha_i, \alpha_j)$ of length $d+1$ consisting of d states α_i for some $i \in [1:I]$ and a final state α_j for some $j \neq i$. Compute the probability $P_i(d) := P[S \mid \text{Model}, s_1 = \alpha_i]$, where the condition $s_1 = \alpha_i$ means that the system is assumed to start with state α_i. From this, compute the expected duration $\overline{d}_i$ for state i, which is defined by $\overline{d}_i := \sum_{d=1}^{\infty} d \cdot P_i(d)$. Finally, determine the expected durations for the states α_1, α_2, and α_3 of the system specified in Figure 5.24.
[**Hint:** Use the fact that $\sum_{d=1}^{\infty} d \cdot a^{d-1} = 1/(1-a)^2$ for a number $a \in [0,1)$.]

Solution to Exercise 5.9. For the probability $P_i(d)$ one has

$$P_i(d) = (a_{ii})^{d-1} \cdot (1 - a_{ii}).$$

From this, one obtains the following formula for the expected duration:

$$\overline{d}_i := \sum_{d=1}^{\infty} d \cdot P_i(d) = (1-a_{ii}) \sum_{d=1}^{\infty} d \cdot (a_{ii})^{d-1} = \frac{1-a_{ii}}{(1-a_{ii})^2} = \frac{1}{1-a_{ii}}.$$

For the system specified in Figure 5.24, one obtains $\overline{d}_1 = 1/(1-0.8) = 5$, $\overline{d}_2 = 1/(1-0.7) \approx 3.33$ and $\overline{d}_3 = 1/(1-0.6) = 2.5$.

Exercise 5.10. Let us consider the HMM as specified in Figure 5.28a. Compute the optimal state sequence and its probability for the observation sequence $O = (\beta_1, \beta_3, \beta_1, \beta_3, \beta_3)$, which is a prefix of the observation sequence used in Figure 5.28b. Compare the result with the one obtained in Figure 5.28b.

Solution to Exercise 5.10. The matrices **D** and **E** are the same as in Figure 5.28b except for deleting the last column, respectively. The probability of the optimal state sequence is $\text{Prob}^* = 0.0033$ (rounded up to four decimal points) with $i_5 = 1$ being the maximizing argument. Starting with this index, backtracking yields the index sequence $(1,1,1,1,1)$ corresponding to the optimal state sequence $S^* = (\alpha_1, \alpha_1, \alpha_1, \alpha_1, \alpha_1)$. Note that this is not simply the prefix of the optimal state sequence in (5.45), even though this holds for the observation sequence.

Exercise 5.11. Let us consider the HMM as specified in Figure 5.28a. Determine the optimal state sequence for the observation sequence $O = (\beta_1, \beta_3^{N-1})$ for each $N \in \mathbb{N}$. Argue why the respective state sequence is optimal.

Solution to Exercise 5.11. One has the following optimal state sequences:

- $N = 1$: $S^* = (\alpha_1)$
- $N = 2$: $S^* = (\alpha_1, \alpha_1)$

- $N = 3$: $S^* = (\alpha_1, \alpha_1, \alpha_1)$
- $N = 4$: $S^* = (\alpha_1, \alpha_3, \alpha_3, \alpha_3)$
- $N > 4$: $S^* = (\alpha_1, \alpha_3^{N-1})$

For $N \in 1,2,3,4$, the optimality follows by applying the Viterbi algorithm from Table 5.2, which yields the following matrices **D** and **E**:

D	$o_1 = \beta_1$
α_1	0.4200
α_2	0.0200
α_3	0

E
α_1
α_2
α_3

$i_1 = 1$

D	$o_1 = \beta_1$	$o_2 = \beta_3$
α_1	0.4200	0.1008
α_2	0.0200	0
α_3	0	0.0336

E	$o_1 = \beta_1$
α_1	1
α_2	1
α_3	1

$i_2 = 1$

D	$o_1 = \beta_1$	$o_2 = \beta_3$	$o_3 = \beta_3$
α_1	0.4200	0.1008	0.0242
α_2	0.0200	0	0
α_3	0	0.0336	0.0161

E	$o_1 = \beta_1$	$o_2 = \beta_3$
α_1	1	1
α_2	1	1
α_3	1	3

$i_3 = 1$

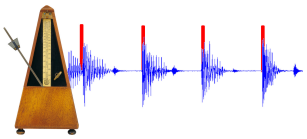
D	$o_1 = \beta_1$	$o_2 = \beta_3$	$o_3 = \beta_3$	$o_4 = \beta_3$
α_1	0.4200	0.1008	0.0242	0.0058
α_2	0.0200	0	0	0
α_3	0	0.0336	0.0161	0.0077

E	$o_1 = \beta_1$	$o_2 = \beta_3$	$o_3 = \beta_3$
α_1	1	1	1
α_2	1	1	3
α_3	1	3	3

$i_4 = 3$

For $N = 4$, the optimal state sequence ends with state α_3. This state yields a higher emission probability for the observation symbol β_3 than the other two states α_1 and α_2. Furthermore, staying in state α_3 yields a higher transition probability than changing to one of the other states α_1 or α_2. In other words, the joint probability of staying in state α_3 and emitting β_3 is higher than changing to another state and emitting β_3. Since the optimal state sequence is $S^* = (\alpha_1, \alpha_3, \alpha_3, \alpha_3)$ for $N = 4$, it follows that the optimal state sequence is $S^* = (\alpha_1, \alpha_3^{N-1})$ for the case $N > 4$.

Chapter 6
Tempo and Beat Tracking



Exercise 6.1. Let $x : \mathbb{Z} \to \mathbb{R}$ be a signal with the nonzero samples $(x(0),\ldots,x(6)) = (0.1,-0.1,0.1,0.9,0.7,0.1,-0.3)$ (all other samples being zero). Furthermore, let $w : \mathbb{Z} \to \mathbb{R}$ be a rectangular window function with nonzero coefficients $w(-1) = w(0) = w(1) = 1$ (i.e., $M = 1$; see Section 6.1.1). Compute all nonzero coefficients of the energy-based novelty function $\Delta_{\text{Energy}} : \mathbb{Z} \to \mathbb{R}$ as defined in (6.3).

Solution to Exercise 6.1. The following table specifies relevant values for the signal x, the squared signal $|x|^2$, the local energy E^x_w (see (6.1)), and the energy-based novelty function Δ_{Energy} (see (6.3)). For each function, only the values for $n \in [-2:8]$ are given (all other values being zero).

n	-2	-1	0	1	2	3	4	5	6	7	8
$x(n)$	0.0	0.0	0.1	-0.1	0.1	0.9	0.7	0.1	-0.3	0.0	0.0
$\lvert x\rvert^2(n)$	0.00	0.00	0.01	0.01	0.01	0.81	0.49	0.01	0.09	0.00	0.00
$\mathrm{E}^x_w(n)$	0.00	0.01	0.02	0.03	0.83	1.31	1.31	0.59	0.10	0.09	0.00
$\Delta_{\text{Energy}}(n)$	0.01	0.01	0.01	0.80	0.48	0.00	0.00	0.00	0.00	0.00	0.00

Exercise 6.2. Let $x : \mathbb{Z} \to \mathbb{R}$ be a discrete signal. Furthermore, let $w : \mathbb{Z} \to \mathbb{R}$ be a rectangular window function of length $2M+1$ centered at time zero, i.e., $w(m) = 1$ for $m \in [-M:M]$ and $w(m) = 0$ otherwise. Then the local energy E^x_w (see (6.1)) is given by

$$\mathrm{E}^x_w(n) := \sum_{m=-M}^{M} x(n+m)^2$$

for $n \in \mathbb{Z}$. In the following, an operation refers to a multiplication, an addition, or a subtraction of two real-valued samples. Determine the overall number of operations that are required to compute $\mathrm{E}^x_w(n)$ for $n \in [0:N-1]$ using a naive approach. Then, describe an improved procedure that reduces the overall number of required operations. How many operations are needed by your procedure?

Solution to Exercise 6.2. Using a naive approach, one needs $2M+1$ multiplications and $2M$ additions for each $n \in [0:N-1]$. Overall, this amounts to $(2M+1)N$ multiplications and $2MN$ additions, thus $4MN+N$ operations in total. First of all, to improve the naive approach, one can precompute and store the values $x(n)^2$ for $n \in [-M:M+N-1]$, which requires $2M+N$ multiplications. Then, one can compute $\mathrm{E}^x_w(0)$ using $2M$ additions. Now, the trick is to derive $\mathrm{E}^x_w(1)$ from $\mathrm{E}^x_w(0)$ by using only one additional subtraction and addition:

$$\mathrm{E}^x_w(1) = \mathrm{E}^x_w(0) - x(-M)^2 + x(1+M)^2.$$

More generally, one can recursively proceed to compute $\mathrm{E}_w^x(n)$ from $\mathrm{E}_w^x(n-1)$ by

$$\mathrm{E}_w^x(n) = \mathrm{E}_w^x(n-1) - x(n-1-M)^2 + x(n+M)^2$$

for $n = 1,\ldots,N-1$, each time using only one subtraction and one addition. Thus, the overall procedure requires $2M+N+2M+2(N-1) = 3N+4M-2$ operations.

Exercise 6.3. Let $\mathcal{Y}$ be an $(N \times (K+1))$ matrix with coefficients $\mathcal{Y}(n,k)$ indexed by $n \in [0:N-1]$ and $k \in [0:K]$. In the following, we consider the matrix $\mathcal{Y}$ defined by

$$\mathcal{Y}^\top = \begin{bmatrix} 0 & 0.1 & 0.1 & 0 & 0.2 & 0.1 \\ 0 & 0 & 0.1 & 0.1 & 0.2 & 0.1 \\ 0 & 0.8 & 0.7 & 0.5 & 0.6 & 0.4 \\ 0 & 0 & 0 & 0 & 0.8 & 0.7 \\ 0 & 0 & 0 & 0.1 & 0 & 0 \end{bmatrix},$$

where $N = 6$ and $K = 4$. (Note that the transposed matrix has been specified.) Interpreting this matrix as a magnitude spectrogram, compute the novelty function $\Delta_{\mathrm{Spectral}}$ as defined in (6.6). Furthermore, compute the local average function μ using $M = 1$ (see (6.7)) and the enhanced novelty function $\bar{\Delta}_{\mathrm{Spectral}}$ (see (6.8)).

Solution to Exercise 6.3. We first compute the values $|\mathcal{Y}(n+1,k) - \mathcal{Y}(n,k)|_{\geq 0}$ for $n \in [0:N-2]$ and $k \in [0:K]$ as in (6.6) yielding an $((N-1) \times (K+1))$ matrix $\mathcal{Y}_{\mathrm{Diff}}$. This matrix is given by

$$(\mathcal{Y}_{\mathrm{Diff}})^\top = \begin{bmatrix} 0.1 & 0 & 0 & 0.2 & 0 \\ 0 & 0.1 & 0 & 0.1 & 0 \\ 0.8 & 0 & 0 & 0.1 & 0 \\ 0 & 0 & 0 & 0.8 & 0 \\ 0 & 0 & 0.1 & 0 & 0 \end{bmatrix}$$

From this one obtains the following values for $\Delta_{\mathrm{Spectral}}$, μ, and $\bar{\Delta}_{\mathrm{Spectral}}$ (using a suitable zeropadding for $\mathcal{Y}_{\mathrm{Diff}}$). For each function, only the values for $n \in [-1:5]$ are given (all other values being zero).

n	-1	0	1	2	3	4	5
$\Delta_{\mathrm{Spectral}}(n)$	0	0.9	0.1	0.1	1.2	0	0
$\mu(n)$	0.300	0.333	0.367	0.467	0.433	0.400	0
$\bar{\Delta}_{\mathrm{Spectral}}(n)$	0	0.567	0	0	0.767	0	0

Exercise 6.4. Realize a bandwise approach for spectral-based novelty detection as outlined at the end of Section 6.1.2. More concretely, let x denote the given music signal sampled at a rate of $F_{\mathrm{s}} = 22050$ Hz and $\mathcal{Y}$ the resulting (possibly compressed) magnitude spectrogram (as in (6.5)) using an STFT window length of $N = 4096$ and a hop size of $H = N/2$. In a first step, divide the frequency range into bands with the first band covering 0–500 Hz, the second 500–1000 Hz, the third 1000–2000 Hz, the fourth 2000–4000 Hz, and the fifth band 4000–11025 Hz. Determine for each of the bands the set of spectral coefficients (similar to (3.3)). Then, compute a novelty

function for each of the bands separately (similar to (6.6)). Finally, compute a single overall novelty function by considering a weighted sum over the bandwise novelty functions using the weighting factor $w_\ell \in \mathbb{R}_{>0}$ for the ℓ^{th} band, $\ell \in [1:5]$. Give a formal description of this procedure by specifying the mathematical details.

Solution to Exercise 6.4. Recall from (2.28) that $F_{\text{coef}}(k) = (k \cdot F_{\text{s}})/N$ is the physical frequency associated with index k of the STFT coefficient $\mathcal{X}(n,k)$, where $k \in [0:K]$ and $K = 2048$. (Recall that $K = N/2$ corresponds to the Nyquist frequency.) Let $F^{\text{low}}_{\text{band}}(\ell)$ be the lower and $F^{\text{up}}_{\text{band}}(\ell)$ be the upper frequency bound (given in Hertz) for the ℓ^{th} frequency band, $\ell \in [1:5]$. Similar to (3.3), we define for each band $\ell \in [1:5]$ the set

$$P(\ell) := \{k \in [1:K] : F^{\text{low}}_{\text{band}}(\ell) \leq F_{\text{coef}}(k) < F^{\text{up}}_{\text{band}}(\ell)\}.$$

Given the parameters $F_{\text{s}} = 22050$ Hz and $N = 4096$, we obtain the following sets:

$$\begin{aligned} P(1) &= \{0,\ldots,92\},\\ P(2) &= \{93,\ldots,185\},\\ P(3) &= \{186,\ldots,371\},\\ P(4) &= \{372,\ldots,743\},\\ P(5) &= \{744,\ldots,2048\}. \end{aligned}$$

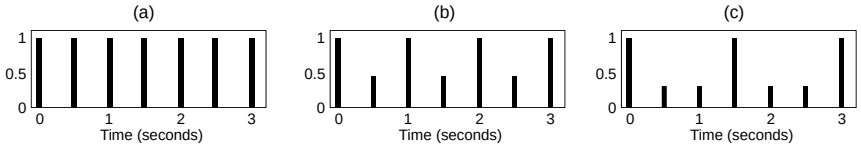
From this, we compute for each band a novelty function $\Delta^{\ell}_{\text{Spectral}}$ by setting

$$\Delta^{\ell}_{\text{Spectral}}(n) := \sum_{k \in P(\ell)} |\mathcal{Y}(n+1,k) - \mathcal{Y}(n,k)|_{\geq 0}$$

for $n \in \mathbb{Z}$. Finally, the overall novelty function $\Delta^{\text{Overall}}_{\text{Spectral}}$ is defined by

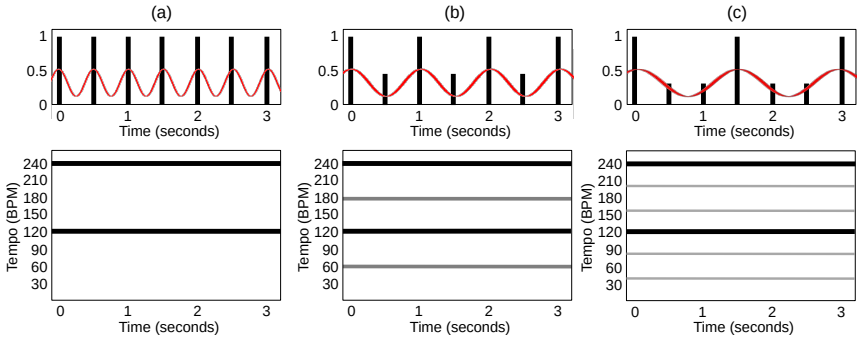
$$\Delta^{\text{Overall}}_{\text{Spectral}}(n) := \sum_{\ell \in [1:5]} w_\ell \cdot \Delta^{\ell}_{\text{Spectral}}(n).$$

Exercise 6.5. In this exercise, we consider the novelty functions corresponding to the click tracks shown in the following figure:



For each of these novelty functions, sketch the Fourier tempogram (see Section 6.2.2) in the tempo range between 20 and 250 BPM. In particular, specify the tempo parameters for which one expects large tempogram coefficients. What is the smallest such parameter (corresponding to the lowest relevant tempo) for each case? Finally, for each of the three novelty functions, indicate visually (as in Figure 6.13) the correlation with the analyzing sinusoid corresponding to this smallest, yet relevant tempo.

Solution to Exercise 6.5.



For the novelty function shown in (a), the only relevant tempo parameters in the range considered correspond to $\tau = 120$ BPM and $\tau = 240$ BPM. The sinusoid shown corresponds to $\tau = 120$ BPM.

In case (b), every second click of the click track has a lower amplitude. Therefore, correlating the novelty function with a sinusoid that corresponds to tempo $\tau = 60$ BPM (half the tempo of $\tau = 120$ BPM) yields a positive coefficient (see also the sinusoid shown in red). This leads to large tempo coefficients for the tempo $\tau = 60$ BPM. However, the size of these coefficients is smaller than the one for the coefficients for the tempo $\tau = 120$ BPM. Furthermore, there are coefficients at the tempo harmonics $\tau = 180$ BPM and $\tau = 240$ BPM.

For the novelty function shown in (c), the lowest, yet relevant tempo is $\tau = 40$ BPM (see also the sinusoid shown in red), which is one third of the tempo of $\tau = 120$ BPM. Furthermore, one may expect relevant coefficients at the harmonics of this tempo value.

Exercise 6.6. Let $x \in \ell^2(\mathbb{Z})$ be a real-valued discrete-time signal. Furthermore, let R_{xx} be the autocorrelation of x, which is given by $R_{xx}(\ell) = \sum_{n\in\mathbb{Z}} x(n)x(n-\ell)$ for each lag parameter $\ell \in \mathbb{Z}$ (see (6.27)). Show that $R_{xx}(0) = \mathrm{E}(x)$ (see (2.41)) and $|R_{xx}(\ell)| \leq R_{xx}(0)$. Furthermore, show that R_{xx} is symmetric, i.e., $R_{xx}(\ell) = R_{xx}(-\ell)$. [**Hint:** Use the Cauchy–Schwarz inequality $|\langle x|y\rangle| \leq \|x\|\|y\|$ from (2.40), which holds for any $x, y \in \ell^2(\mathbb{Z})$.]

Solution to Exercise 6.6. By definition, one obtains

$$R_{xx}(0) = \sum_{n\in\mathbb{Z}} x(n)x(n) = \sum_{n\in\mathbb{Z}} |x(n)|^2 = \mathrm{E}(x).$$

For a given $\ell \in \mathbb{Z}$, define the signal y by setting $y(n) := x(n-\ell)$, $n \in \mathbb{Z}$. Then, $R_{xx}(\ell) = \langle x|y\rangle$ and $\|y\| = \|x\|$. Furthermore, by the Cauchy–Schwarz inequality, one obtains

$$|R_{xx}(\ell)| = |\langle x|y\rangle| \leq \|x\|\|y\| = \|x\|^2 = R_{xx}(0).$$

Finally, the symmetry follows from

$$R_{xx}(\ell) = \sum_{n\in\mathbb{Z}} x(n)x(n-\ell) = \sum_{n\in\mathbb{Z}} x(n+\ell)x(n+\ell-\ell) = \sum_{n\in\mathbb{Z}} x(n)x(n+\ell) = R_{xx}(-\ell),$$

where we have exploited the fact that one can apply an index shift when summing over the integers.

Exercise 6.7. Let $x : \mathbb{Z} \to \mathbb{R}$ be a real-valued signal. Assume that the support of x lies in the interval $[-M : M]$ for some $M \in \mathbb{N}$. Let R_{xx} be the autocorrelation as defined in (6.27). Show that $R_{xx}(\ell) = 0$ for $|\ell| \geq 2M+1$. Furthermore, show that at most $2M+1-|\ell|$ of the summands in (6.29) are nonzero.

Solution to Exercise 6.7. As for the number of summands, note that in the sum of $R_{xx}(\ell) = \sum_{m\in\mathbb{Z}} x(m)x(m-\ell)$, the first factor $x(m)$ is possibly nonzero only for $m \in [-M : M]$ and the second factor $x(m-\ell)$ is possibly nonzero only for $m \in [-M+\ell : M+\ell]$. For $\ell \geq 0$, one obtains

$$[-M : M] \cap [-M+\ell : M+\ell] = [-M+\ell : M].$$

Note that $[-M+\ell : M] = \emptyset$ for $\ell \geq 2M+1$. For $\ell \leq 0$, one obtains

$$[-M : M] \cap [-M+\ell : M+\ell] = [-M : M+\ell].$$

Note that $[-M : M+\ell] = \emptyset$ for $\ell \leq -(2M+1)$. This implies that the number of nonzero summands is at most

$$\big|[-M : M] \cap [-M+\ell : M+\ell]\big| = \max(2M+1-|\ell|, 0).$$

In particular, the number of nonzero summands is zero in the case that $|\ell| \geq 2M+1$. This shows that $R_{xx}(\ell) = 0$ for $|\ell| \geq 2M+1$.

Exercise 6.8. Let $\Delta : \mathbb{Z} \to \mathbb{R}$ be a novelty function with a feature rate of 10 Hz. Furthermore, let $\mathcal{T}^{\mathrm{A}}$ be the autocorrelation tempogram derived from Δ (see (6.31)). What is the maximal tempo that is captured by $\mathcal{T}^{\mathrm{A}}$?

Solution to Exercise 6.8. Having a feature rate of 10 Hz, each time frame of the resulting time-lag representation corresponds to $r = 0.1$ sec. Let $\tau^{\max}$ denote the maximal tempo that is captured by $\mathcal{T}^{\mathrm{A}}$. This tempo corresponds to time lag $\ell = 1$ (given in frames). By (6.30), one obtains

$$\tau^{\max} = \frac{60}{r \cdot \ell} = 600 \text{ BPM}.$$

Exercise 6.9. In this exercise, we consider a discrete cyclic tempogram representation $\mathcal{C}_{\tau_0}$ using a reference tempo $\tau_0 = 60$ BPM (see (6.35)). For computing $\mathcal{C}_{\tau_0}$, we use four tempo octaves ranging from $\tau = 30$ to $\tau = 480$ BPM, where each octave is logarithmically sampled using $M \in \mathbb{N}$ tempo parameters. Specify a formula for the tempo values that are needed to compute $\mathcal{C}_{\tau_0}$. Furthermore, using $M = 10$, determine the eleven tempo values between $\tau = 60$ and $\tau = 120$ BPM. Next, assume that $\mathcal{C}_{\tau_0}$ is derived from an autocorrelation tempogram based on a feature rate of 10 Hz. Determine the lag parameters corresponding to the eleven computed tempo values. Which problems arise? Make suggestions to alleviate these problems.

Solution to Exercise 6.9. The four tempo octaves starting with 30 BPM can be logarithmically sampled with $M \in \mathbb{N}$ parameters per octave via

$$2^{m/M} \cdot 30 \text{ BPM}$$

for $m = [0:4M]$. Thus, using $M = 10$, one obtains the following eleven (rounded) tempo values (given in BPM) for $m \in [10:20]$:

$$60.0,\ 64.3,\ 68.9,\ 73.9,\ 79.2,\ 84.9,\ 90.9,\ 97.5,\ 104.5,\ 112.0,\ 120.0.$$

Having a feature rate of 10 Hz, each time frame of the resulting time-lag representation corresponds to $r = 0.1$ sec. From (6.30), one obtains the formula $\ell = 60/(r \cdot \tau)$ for a given tempo value τ. This yields the following eleven lag values:

$$10.0,\ 9.3,\ 8.7,\ 8.1,\ 7.6,\ 7.1,\ 6.6,\ 6.2,\ 5.7,\ 5.3,\ 5.0.$$

Since a lag parameter is an integer, these values need to be further quantized or rounded. This leads to inaccuracies in the tempo values to be considered. In particular, for small lags (high tempi), the resolution becomes poor. To alleviate this problem, one may increase the feature rate of the novelty function. This not only increases the feature rate of the resulting autocorrelation tempogram, but also the resolution of the lag axis.

Exercise 6.10. For a given parameter $N \in \mathbb{N}$, let $\mathcal{B}^N$ be the space of all possible beat sequences within the interval $[1:N]$ (see Section 6.3.2). Determine the number $|\mathcal{B}^N|$. Furthermore, given a length parameter $L \in [0:N]$, determine the number of beat sequences of length L. Finally, let $\mathcal{B}^N_n \subset \mathcal{B}^N$ denote the subset of all beat sequences that end in $n \in [0:N]$ (where the case $n = 0$ refers to the empty beat sequence). Determine the number $|\mathcal{B}^N_n|$. Finally, show that $\mathcal{B}^N = \cup_{n\in[0:N]} \mathcal{B}^N_n$ (see (6.44)).

Solution to Exercise 6.10. The beat sequences in $[1:N]$ correspond to subsets of $[1:N]$. Therefore, the number of beat sequences equals the number of possible subsets of $[1:N]$, which is 2^N. Similarly, the number of beat sequences of length L corresponds to the number of subsets of $[1:N]$ having size L. This number is $\binom{N}{L}$. Note that $\sum_{L=0}^{N} \binom{N}{L} = 2^N$.

The beat sequence ending with $n = 0$ is, by definition, the empty beat sequence. Next, the beat sequences ending in some $n \in [1:N]$ are in a one-to-one correspondence to subsets of $[1:n-1]$. Therefore, $|\mathcal{B}^N_n| = 2^{n-1}$. Obviously, a beat sequence $(b_1,\ldots,b_L) \in \mathcal{B}^N$ is contained in $\mathcal{B}^N_n$ with $n = b_L$. This shows that $\mathcal{B}^N = \cup_{n\in[0:N]} \mathcal{B}^N_n$. Note that the sets $\mathcal{B}^N_n$ are pairwise disjoint, i.e., $\mathcal{B}^N_n \cap \mathcal{B}^N_m = \emptyset$ for $n,m \in [0:N]$ with $n \neq m$. This is also reflected by the fact that

$$\sum_{n=0}^{N} |\mathcal{B}^N_n| = 1 + \sum_{n=1}^{N} 2^{n-1} = 1 + 2^N - 1 = 2^N.$$

Exercise 6.11. Given a novelty function $\Delta : [1:N] \to \mathbb{R}$, analyze the computational complexity of the beat tracking procedure described in Section 6.3.2 (see also Table 6.1) in terms of memory requirements as well as in terms of the number of required operations. Assume that an operation is an addition, a multiplication, an evaluation of $P_{\hat{\delta}}$, or a maximization (where maximization over a set of $M \in \mathbb{N}$ elements counts as M operations).

Solution to Exercise 6.11. As for the memory requirements, one essentially needs to store the novelty function Δ, the accumulated score values $\mathbf{D}$, the predecessor information $\mathbf{P}$, and the beat sequence B^*. This requires $O(N)$.

The number of operations is dominated by the recursion (6.47), which is applied for $n = 1, \ldots, N$. In the n^{th} step, one needs to maximize over a set consisting of $n-1$ elements and perform n additions, $n-1$ multiplications, and $n-1$ evaluations of $P_{\hat{\delta}}$. This leads to an overall number of operations on the order of $O(\sum_{n=1}^{N} n) = O(N^2)$.

Exercise 6.12. Apply the beat tracking procedure described in Section 6.3.2 (see also Table 6.1) to the novelty function $\Delta : [1:N] \to \mathbb{R}$ with $N = 11$ given by the following values:

n	1	2	3	4	5	6	7	8	9	10	11
$\Delta(n)$	0.1	0.0	1.0	0.0	1.0	0.8	0.0	0.2	0.4	1.0	0.0

For the computations, use the weight parameter $\lambda = 1$ and the following values for the penalty function $P_{\hat{\delta}}$ which favors the beat period $\hat{\delta} = 3$ (note that, for the sake of simplicity, these values are not obtained from (6.40)):

n	1	2	3	4	5	6	7	8	9	10	11
$P_{\hat{\delta}}(n)$	-2	-0.2	1.0	0.5	-0.1	-1	-1.5	-3	-5	-8	-12

Compute the accumulated score values $\mathbf{D}(n)$ and the predecessors $\mathbf{P}(n)$ for $n \in [1:N]$. Furthermore, derive the optimal beat sequence B^*.

Solution to Exercise 6.12. Applying the recursion (6.48), one obtains the following values:

n	1	2	3	4	5	6	7	8	9	10	11
$\Delta(n)$	0.1	0.0	1.0	0.0	1.0	0.8	0.0	0.2	0.4	1.0	0.0
$\mathbf{D}(n)$	0.1	0.0	1.0	1.1	2.0	2.8	2.1	3.2	4.2	4.3	4.2
$\mathbf{P}(n)$	0	0	0	1	2	3	4	5	6	6	8

The maximizing index for $\mathbf{D}$ is $n^* = 10$, which determines the last beat b_L of the optimal beat sequences. Backtracking through $\mathbf{P}$ yields $\mathbf{P}(10) = 6$, $\mathbf{P}(6) = 3$, and $\mathbf{P}(3) = 0$. This yields $B^* = (b_1, b_2, b_3) = (3, 6, 10)$ and $L = 3$.

Exercise 6.13. The penalty function $P_{\hat{\delta}}$ defined in (6.40) (see also Figure 6.21) decreases rapidly with larger deviations from the ideal beat period $\hat{\delta}$. Therefore, it becomes unlikely that the predecessor m of some beat position n lies far from the position $n - \hat{\delta}$. This observation can be used to achieve significant savings by restricting the search space $m \in [1 : n-1]$ in the maximization (6.47). For example, assuming that the next beat to be estimated has at least the distance $\hat{\delta}/2$ and at

most the distance $2\hat{\delta}$ from its predecessor beat, one may replace the search space $m \in [1:n-1]$ by the constrained search space $m \in [1:n-1] \cap [n-2\hat{\delta}, n-\hat{\delta}/2]$. Analyze the computational complexity of the modified procedure (as in Exercise 6.11). Compare the result with the original procedure.

Solution to Exercise 6.13. As in the original approach, the number of operations is dominated by the recursion (6.47), which is applied for $n = 1, \ldots, N$. Now, in the n^{th} step of the constrained procedure, one considers $[1:n-1] \cap [n-2\hat{\delta}, n-\hat{\delta}/2]$ instead of $[1:n-1]$. Thus, in each step, the number of required operations is at most linear in the size of the interval $[n-2\hat{\delta}, n-\hat{\delta}/2]$, which is $O(\hat{\delta})$ (instead of $O(n)$ as in the original approach). This leads to an overall number of operations on the order of $O(\hat{\delta} \cdot N) = O(N)$, since $\hat{\delta}$ is a constant independent of N. Thus, the overall complexity is linear in N (as opposed to being quadratic in N as for the original approach).

Exercise 6.14. Recall that a beat sequence $B = (b_1, b_2, \ldots, b_L)$ is a sequence of increasing indices $b_\ell \in [1:N]$, $\ell \in [1:L]$. Mathematically, this is identical to the notion of a boundary annotation, which we introduced for evaluating novelty-based segmentation procedures in the context of music structure analysis (see Section 4.5.4). Therefore, to evaluate a beat tracking procedure, one can use exactly the same evaluation measures as for novelty detection. Following Section 4.5.4, let B^{Ref} be a reference beat sequence and B^{Est} an estimated beat sequence. Furthermore, let $\tau \geq 0$ be a tolerance parameter for the maximal acceptable deviation. Similar to (4.57), an estimated beat $b^{\text{Est}} \in B^{\text{Est}}$ is considered **correct** if it lies within the τ-neighborhood of a reference beat $b^{\text{Ref}} \in B^{\text{Ref}}$:

$$|b^{\text{Est}} - b^{\text{Ref}}| \leq \tau.$$

Following Section 4.5.4, introduce the notions of true positives, false positives, and false negatives, and then derive the precision, recall, and F-measure. Furthermore, using $\tau = 1$, compute these measures for the following beat sequences:

$$B^{\text{Ref}} = (10, 20, 30, 40, 50, 60, 70, 80, 90)$$
$$B^{\text{Est}} = (10, 19, 26, 34, 42, 50, 61, 70, 78, 89)$$

Solution to Exercise 6.14. The true positives (TP) are defined to be the beats $b^{\text{Est}} \in B^{\text{Est}}$ that are correct, and the false positives (FP) are the beats $b^{\text{Est}} \in B^{\text{Est}}$ that are not correct. Furthermore, the false negatives (FN) are defined to be the beats $b^{\text{Ref}} \in B^{\text{Ref}}$ with no estimated beat in a τ-neighborhood. For the example, the true positives are $\{10, 19, 50, 61, 70, 89\}$, the false positives are $\{26, 34, 42, 78\}$, and the false negatives are $\{30, 40, 80\}$. Based on these definitions, one obtains the following precision, recall, and F-measure:

$$\begin{aligned}
\mathrm{P} &= \#\mathrm{TP}/(\#\mathrm{TP} + \#\mathrm{FP}) = 6/10 = 3/5, \\
\mathrm{R} &= \#\mathrm{TP}/(\#\mathrm{TP} + \#\mathrm{FN}) = 6/9 = 2/3, \\
\mathrm{F} &= 2\mathrm{PR}/(\mathrm{P} + \mathrm{R}) = (2 \cdot (3/5) \cdot (2/3))/(3/5 + 2/3) = 12/19.
\end{aligned}$$

Exercise 6.15. In the evaluation measure considered in Exercise 6.14, the beat positions were evaluated independently of each other. However, when tapping to the beat of music, a listener obviously requires the temporal context of several consecutive beats. Therefore, in evaluating beat tracking procedures, it seems natural to consider beats in the temporal context instead of looking at the beat positions individually. To account for these temporal dependencies, we now introduce a context-sensitive evaluation measure. Let $B^{\mathrm{Ref}} = (r_1, r_2, \ldots, r_M)$ be a reference beat sequence with $r_m \in [1:N]$, $m \in [1:M]$. Similarly, let $B^{\mathrm{Est}} = (b_1, b_2, \ldots, b_L)$ be an estimated beat sequence with $b_\ell \in [1:N]$, $\ell \in [1:L]$. Furthermore, let $K \in \mathbb{N}$ be a parameter that specifies the temporal context measured in beats, and let $\tau \geq 0$ be a tolerance parameter for the maximal acceptable deviation. Then, an estimated beat b_ℓ is considered a **K-correct detection** if there exists a subsequence $b_i, \ldots, b_{i+K-1}$ of B^{Est} containing b_ℓ (i.e., $\ell \in [i : i+K-1]$) as well as a subsequence $r_j, \ldots, r_{j+K-1}$ of B^{Ref} such that

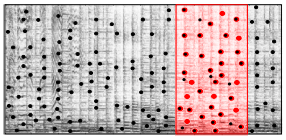
$$|b_{i+k} - r_{j+k}| \leq \tau$$

for all $k \in [0 : K-1]$. Intuitively, for a beat to be considered K-correct, one requires an entire track consisting of K consecutive estimated beats that match (up to the tolerance τ) a track of K consecutive reference beats. Note that a single outlier in the estimated beats voids this property. Let L_K be the number of K-correct estimated beats. Then, we define the context-sensitive precision $\mathrm{P}_K := L_K/L$, recall $\mathrm{R}_K := L_K/M$, and F-measure $\mathrm{F}_K := 2\mathrm{P}_K\mathrm{R}_K/(\mathrm{P}_K + \mathrm{R}_K)$. For B^{Ref} and B^{Est} as specified in Exercise 6.14, determine the set of K-correct beat sequences as well as the context-sensitive precision, recall, and F-measure for $\tau = 1$ and $K \in \{1,2,3,4\}$.

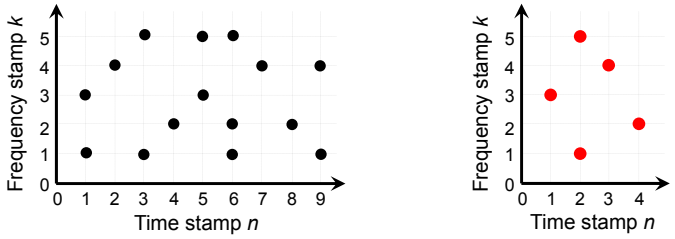
Solution to Exercise 6.15. In the following table, the K-correct estimated beats are indicated by the symbol '$\times$.' Furthermore, P_K, R_K, and F_K are shown in the last three columns for $K \in \{1,2,3,4\}$.

	B^{Ref}	10	20		30	40	50	60	70	80	90			
K	B^{Est}	10	19	26	34	42	50	61	70	78	89	P_K	R_K	F_K
1	1-correct	$\times$	$\times$				$\times$	$\times$	$\times$		$\times$	$3/5$	$2/3$	$12/19$
2	2-correct	$\times$	$\times$				$\times$	$\times$	$\times$			$1/2$	$5/9$	$10/19$
3	3-correct						$\times$	$\times$	$\times$			$3/10$	$1/3$	$6/19$
4	4-correct											0	0	0

Chapter 7
Content-Based Audio Retrieval



Exercise 7.1. Consider the constellation maps $\mathcal{C}(\mathcal{D})$ (left) and $\mathcal{C}(\mathcal{Q})$ (right) as specified by the figure below. Determine the resulting matching function $\Delta_{\mathcal{C}} : \mathbb{Z} \to \mathbb{N}_0$ as defined in (7.3) by shifting $\mathcal{C}(\mathcal{Q})$ over $\mathcal{C}(\mathcal{D})$ (see Figure 7.5).



Solution to Exercise 7.1. The values $\Delta_{\mathcal{C}}(m)$ for the shift indices $m \in [-2:8]$ of the matching function $\Delta_{\mathcal{C}}$ are given by the following table. The values for all other shift indices are zero.

	Shift index *m*										
	-2	-1	0	1	2	3	4	5	6	7	8
Matching function	**0**	**2**	**2**	**2**	**1**	**1**	**5**	**0**	**1**	**1**	**0**

Exercise 7.2. Let $\mathcal{F}(\mathcal{D}) := \mathcal{C}(\mathcal{D})$ and $\mathcal{F}(\mathcal{Q}) := \mathcal{C}(\mathcal{Q})$ be specified as in the figure of Exercise 7.1. Determine the inverted lists and the indicator functions as in Figure 7.6. Then compute the matching function $\Delta_{\mathcal{F}}$ as in (7.8).

Solution to Exercise 7.2. The following figure shows the five inverted lists for $\mathcal{F}(\mathcal{D})$ (left) and illustrates the computation of the matching function $\Delta_{\mathcal{F}}$ using the indicator functions of the suitably shifted inverted lists (right):

L(5) = (3,5,6)
L(4) = (2,7,9)
L(3) = (1,5)
L(2) = (4,6,8)
L(1) = (1,3,6,9)

Query (*n*,*h*)	*L*(*h*) - *n*	Indicator functions										
		…	-1	0	1	2	3	4	5	6	7	…
(1,3)	(0,4)	0	0	1	0	0	0	1	0	0	0	0
(2,1)	(-1,1,4,7)	0	1	0	1	0	0	1	0	0	1	0
(2,5)	(1,3,4)	0	0	0	1	0	1	1	0	0	0	0
(3,4)	(-1,4,6)	0	1	0	0	0	0	1	0	1	0	0
(4,2)	(0,2,4)	0	0	1	0	1	0	1	0	0	0	0
Matching function		**0**	**2**	**2**	**2**	**1**	**1**	**5**	**0**	**1**	**1**	**0**

Exercise 7.3. In this exercise, we look at the survival probability of a hash that consists of two frequency stamps and a time stamp difference (see Section 7.1.4). Let $p \in [0,1]$ be the probability of a spectral peak surviving in the query audio fragment, and let $F \in \mathbb{N}$ denote the fan-out of the target zone. Assuming that the peak survival probability is independent and identically distributed, show that the joint probability of the anchor point and at least one target point in its target zone surviving is given by (7.16):

$$p \cdot (1-(1-p)^F).$$

Furthermore, compute the number $(1-(1-p)^F)$ for $p \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ in combination with different $F \in \{1, 5, 10, 20, 40\}$. Discuss the results and the kind of trade-offs involved.

Solution to Exercise 7.3. The probability that one peak survives is p. Now, let us consider a set of F peaks (the ones contained in a target zone). Assuming that the peak survival probability is independent and identically distributed, the probability that none of those peaks survive is $(1-p)^F$. Therefore, the probability that at least one of these peaks survives is $1-(1-p)^F$. Therefore, the probability that at least one of these peaks survives, combined with the anchor peak surviving is $p \cdot (1-(1-p)^F)$. This proves (7.16). The values $(1-(1-p)^F)$ for $p \in \{0.1, 0.2, 0.3, 0.4, 0.5\}$ and $F \in \{1, 5, 10, 20, 40\}$ are as follows (rounded to four decimal places):

	$p=0.1$	$p=0.2$	$p=0.3$	$p=0.4$	$p=0.5$
$F=1$	0.1000	0.2000	0.3000	0.4000	0.5000
$F=5$	0.4095	0.6723	0.8319	0.9222	0.9688
$F=10$	0.6513	0.8926	0.9718	0.9940	0.9990
$F=20$	0.8784	0.9885	0.9992	1.0000	1.0000
$F=40$	0.9852	0.9999	1.0000	1.0000	1.0000

The survival probability of at least one target point surviving increases drastically by increasing F. However, increasing the fan-out F also increases the storage requirements on the database side (which depends linearly on F) and reduces the effect on the speed up (which depends on F^2 in a reciprocal fashion); see also (7.15).

Exercise 7.4. Let $\mathcal{F} = \mathbb{R}$ be a feature space and $c : \mathcal{F} \times \mathcal{F} \to \mathbb{R}_{\geq 0}$ be a local cost measure defined by $c(x,y) = |x-y|$ for $x, y \in \mathbb{R}$ (see also Exercise 3.10). Given the sequences $X = (x_1, \ldots, x_N) = (3,0,6)$ of length $N = 3$ and $Y = (y_1, \ldots, y_M) = (2,4,0,4,0,0,5,2)$ of length $M = 8$, compute the matching function $\Delta_{\text{Diag}} : [0 : M-N] \to \mathbb{R}$ (see (7.20)) as well as the resulting best match (see (7.23)). Furthermore, compute the DTW-based matching function $\Delta_{\text{DTW}} : [1 : M] \to \mathbb{R}$ using the step size set $\Sigma = \{(1,0), (0,1), (1,1)\}$ (see (7.29)) as well as the resulting optimal subsequence $Y(a^* : b^*)$ (see (7.25)).

Solution to Exercise 7.4. The following figure shows the cost matrix **C** as well the accumluated cost matrix **D**:

C

6	4	0	6	2	6	6	1	4
0	2	6	0	4	0	0	5	2
3	1	3	3	1	3	3	2	1
	2	4	0	4	0	0	5	2

D

6	7	3	7	3	7	7	2	6
0	3	7	1	5	1	1	6	3
3	1	1	3	1	3	3	2	1
	2	4	0	4	0	0	5	2

From **C**, one obtains

$$\Delta_{\mathrm{Diag}}(0:5) = \frac{1}{3}(13,5,13,7,4,12).$$

The index $m^* \in [0 : M-N]$ that minimizes the matching function is $m^* = 4$. The best match is $Y(1+m^* : N+m^*) = Y(5:7) = (0,0,5)$ (see (7.23)). From **D**, one obtains

$$\Delta_{\mathrm{DTW}}(1:8) = \frac{1}{3}(7,3,7,3,7,7,2,6),$$

which yields $b^* = 7$. By backtracking, one obtains $a^* = 4$. Thus the optimal subsequence is $Y(4:7) = (4,0,0,5)$.

Exercise 7.5. In this exercise, we show how the matching procedures of Section 7.2 can be applied to a concatenated feature sequence of different recordings, while avoiding matches across different recordings. As in Section 7.2.2, let $X = (x_1,\ldots,x_N)$ be a feature sequence of a query audio fragment. Furthermore, let $Y^i = (y^i_1,\ldots,y^i_{M_i})$ be feature sequences of length $M_i \geq N$ of two database recordings indexed by $i \in \{1,2\}$. Let Δ^i_{Diag} be the two matching functions obtained by comparing X and Y^i for $i \in \{1,2\}$ (see (7.20)). Next, we concatenate both feature sequences by defining

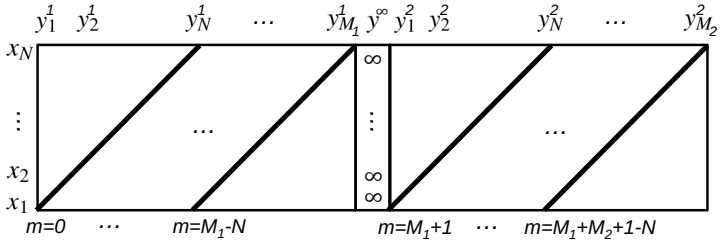
$$Y := (y^1_1,\ldots,y^1_{M_1},y^\infty,y^2_1,\ldots,y^2_{M_2}),$$

where y^∞ denotes a feature vector consisting of ∞ entries. Assume that $c(x,y^\infty) := \infty$ for any feature vector x. Furthermore, assume that the sum of the value ∞ with a finite value is defined to be ∞ and that the minimum over a set containing finite values as well as the value ∞ is defined to be the minimum over the finite values (see also Exercise 3.13). Using these calculation rules, let Δ_{Diag} be the matching function obtained by comparing X and Y. Describe the relation between Δ_{Diag}, Δ^1_{Diag}, and Δ^2_{Diag}. What happens in the case that Y is simply defined as the concatenation of Y^1 and Y^2 (without the additional y^∞ vector)?

Similarly, define the matching functions Δ_{DTW}, Δ^1_{DTW}, and Δ^2_{DTW} based on the step size set $\Sigma = \{(1,0),(0,1),(1,1)\}$ (see (7.29)) and discuss their relations. What happens when the step size condition $\Sigma = \{(2,1),(1,2),(1,1)\}$ is used? Describe a strategy to avoid matches across different recordings in this setting.

Solution to Exercise 7.5. Let $M = M_1 + M_2 + 1$ be the length of the concatenated sequence Y. Recall that the matching function $\Delta_{\mathrm{Diag}} : [0 : M-N] \to \mathbb{R}$ is defined by $\Delta_{\mathrm{Diag}}(m) := \frac{1}{N}\sum_{n=1}^{N} c(x_n, y_{n+m})$ for $m \in [0 : M-N]$ (see (7.20)). Similarly, one

obtains the matching functions $\Delta^i_{\mathrm{Diag}} : [0 : M_i - N] \to \mathbb{R}$ for $i \in \{1,2\}$. The following figure illustrates the cost matrix $\mathbf{C}$ and the computation of Δ_{Diag}:



From this, one can see that

$$\Delta_{\mathrm{Diag}}(m) := \begin{cases} \Delta^1_{\mathrm{Diag}}(m) & \text{for } m \in [0 : M_1 - N], \\ \infty & \text{for } m \in [M_1 - N + 1 : M_1], \\ \Delta^2_{\mathrm{Diag}}(m - M_1 - 1) & \text{for } m \in [M_1 + 1 : M - N]. \end{cases}$$

Without the feature y^∞, one would obtain matches across Y^1 and Y^2 with a cost less than ∞, which may lead to false positives. In the case of the matching functions $\Delta_{\mathrm{DTW}} : [1 : M] \to \mathbb{R}$, $\Delta^1_{\mathrm{DTW}} : [1 : M_1] \to \mathbb{R}$, and $\Delta^2_{\mathrm{DTW}} : [1 : M_2] \to \mathbb{R}$ based on the step size set $\Sigma = \{(1,0),(0,1),(1,1)\}$, one obtains

$$\Delta_{\mathrm{DTW}}(m) := \begin{cases} \Delta^1_{\mathrm{DTW}}(m) & \text{for } m \in [1 : M_1], \\ \infty & \text{for } m = M_1 + 1, \\ \Delta^2_{\mathrm{Diag}}(m - M_1 - 1) & \text{for } m \in [M_1 + 2 : M]. \end{cases}$$

This follows by a similar argumentation as the one in Exercise 3.13. In the case of the step size set $\Sigma = \{(2,1),(1,2),(1,1)\}$, the feature y^∞ may be skipped by going from the cell (n, M_1) (matching x_n with $y^1_{M_1}$) to the cell $(n+1, M_1+2)$ (matching x_{n+1} with y^2_1) using the step $(1,2)$. This leads to matches across Y^1 and Y^2 with a cost less than ∞. To avoid such matches, one needs to include the feature y^∞ twice in the concatenation:

$$Y := (y^1_1, \ldots, y^1_{M_1}, y^\infty, y^\infty, y^2_1, \ldots, y^2_{M_2}).$$

Exercise 7.6. Let $X = (x_1, x_2, \ldots, x_N)$ and $Y = (y_1, y_2, \ldots, y_M)$ be two feature sequences over the feature space $\mathcal{F}$, and let $c : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$ be a local cost measure. The task of **subsequence DTW** is to determine the subsequence of Y that best matches the sequence X. This subsequence is given by

$$(a^*, b^*) := \operatorname*{argmin}_{(a,b):1 \le a \le b \le M} \mathrm{DTW}\big(X, Y(a:b)\big)$$

(see (7.25)). Following Section 7.2.3, specify the subsequence DTW algorithm (using the step size set $\Sigma = \{(1,0),(0,1),(1,1)\}$) similar to Table 3.2. Given the cost

matrix $\mathbf{C}$, the algorithm should output the accumulated cost matrix $\mathbf{D}$, the indices $a^*, b^* \in [1:M]$, as well as an optimal warping path between X and $Y(a^*:b^*)$.

Solution to Exercise 7.6. The following table shows a specification of the algorithm for subsequence DTW:

Algorithm: SUBSEQUENCE DTW

Input: Cost matrix $\mathbf{C}$ of size $N \times M$
Output: Accumulated cost matrix $\mathbf{D}$
Indices $a^*, b^* \in [1:M]$ of an optimal subsequence of Y
Optimal warping path P^* between X and $Y(a^*:b^*)$

Procedure: Initialize $(N \times M)$ matrix $\mathbf{D}$ by $\mathbf{D}(n,1) = \sum_{k=1}^{n} \mathbf{C}(k,1)$ for $n \in [1:N]$ and $\mathbf{D}(1,m) = \mathbf{C}(1,m)$ for $m \in [1:M]$. Then compute in a nested loop for $n = 2,\ldots,N$ and $m = 2,\ldots,M$:

$$\mathbf{D}(n,m) = \mathbf{C}(n,m) + \min\{\mathbf{D}(n-1,m-1), \mathbf{D}(n-1,m), \mathbf{D}(n,m-1)\}.$$

Set $b^* = \operatorname{argmin}_{b \in [1:M]} \mathbf{D}(N,b)$. (If 'argmin' is not unique, take smallest index.)
Set $\ell = 1$ and $q_\ell = (N, b^*)$.
Then repeat the following steps until $q_\ell = (1,m)$ for some $m \in [1:M]$:

Increase ℓ by one and let $(n,m) = q_{\ell-1}$.
If $m = 1$, then $q_\ell = (n-1, 1)$,
else $q_\ell = \operatorname{argmin}\{\mathbf{D}(n-1,m-1), \mathbf{D}(n-1,m), \mathbf{D}(n,m-1)\}$.
(If 'argmin' is not unique, take lexicographically smallest cell.)

Set $L = \ell$ and $a^* = m$. Return $\mathbf{D}$, a^*, b^*, and $P^* = (q_L, q_{L-1}, \ldots, q_1)$.

Exercise 7.7. The goal of this exercise is to show how diagonal matching is related to DTW-based matching. Let $X = (x_1, x_2, \ldots, x_N)$ and $Y = (y_1, y_2, \ldots, y_M)$ be two sequences, and let Δ_{Diag} be the matching function based on diagonal matching (see Section 7.2.2). Furthermore, let Δ_{DTW} be the DTW-based matching function using the step size set $\Sigma = \{(1,1)\}$ (instead of using $\Sigma = \{(1,0),(0,1),(1,1)\}$ as in Section 7.2.3). First, describe how the DTW-based procedure needs to be modified when using $\Sigma = \{(1,1)\}$. Then, explain how Δ_{Diag} and Δ_{DTW} are related.

Solution to Exercise 7.7. When using the step size set $\Sigma = \{(1,1)\}$, one can initialize the DTW-based procedure in a fashion similar to Exercise 3.13. In the subsequence DTW case, one extends the accumulated cost matrix $\mathbf{D}$ by an additional column indexed by 0 and then defines $\mathbf{D}(n,0) := \infty$ for $n \in [1:N]$. The first row of $\mathbf{D}$ is initialized as before by setting $\mathbf{D}(1,m) := \mathbf{C}(1,m)$ for $m \in [1:M]$ (see (7.27)). In the recursion, the step size set $\Sigma = \{(1,1)\}$ enforces that only diagonal steps are allowed. As a result, this procedure yields a matching function $\Delta_{\text{DTW}} : [1:M] \to \mathbb{R}$ with $\Delta_{\text{DTW}}(m) = \infty$ for $m \in [1:N-1]$. Furthermore, one obtains

$$\Delta_{\text{DTW}}(m) = \frac{1}{N}\,\mathbf{D}(N,m) = \frac{1}{N}\sum_{n=1}^{N} c(x_n, y_{m-N+n})$$

for $m \in [N:M]$ (see also (7.29)). Recall from (7.20) that diagonal matching yields a function $\Delta_{\text{Diag}} : [0:M-N] \to \mathbb{R}$ with

$$\Delta_{\mathrm{Diag}}(m) = \frac{1}{N}\sum_{n=1}^{N} c(x_n, y_{n+m})$$

for $m \in [1 : M-N]$ (see (7.20)). Therefore, $\Delta_{\mathrm{Diag}}(m) = \Delta_{\mathrm{DTW}}(m+N)$ for $m \in [0 : M-N]$.

Intuitively speaking, in diagonal matching, the subsequences of Y are compared with X in a "forward" manner, where the index m in $\Delta_{\mathrm{Diag}}(m)$ indicates the beginning of the considered matching subsequence. In contrast, in DTW-based matching, the subsequences of Y are compared with X in a "backward" manner, where the index m in $\Delta_{\mathrm{DTW}}(m)$ indicates the end of the considered matching subsequence.

Exercise 7.8. For a sequence $S = (s_1, \ldots, s_L)$, let $\mathrm{Rev}(S) = (r_1, \ldots, r_L)$ with $r_\ell := s_{L-\ell+1}$, $\ell \in [1 : L]$ denote the reversed sequence. Now, let $X = (x_1, x_2, \ldots, x_N)$ and $Y = (y_1, y_2, \ldots, y_M)$ be two feature sequences as in Section 7.2.3. Using the step size set $\Sigma = \{(1,0), (0,1), (1,1)\}$, let $\Delta_{\mathrm{DTW}}[X,Y]$ be the DTW-based matching function for X and Y and $\Delta_{\mathrm{DTW}}[\mathrm{Rev}(X), \mathrm{Rev}(Y)]$ be the one for $\mathrm{Rev}(X)$ and $\mathrm{Rev}(Y)$. Assume that the indices

$$(a^*, b^*) := \operatorname*{argmin}_{(a,b):1\le a\le b\le M} \mathrm{DTW}\big(X, Y(a:b)\big)$$

(see (7.25)) are uniquely determined. In Section 7.2.3, we showed that

$$b^* = \operatorname*{argmin}_{m\in[1:M]} \Delta_{\mathrm{DTW}}[X,Y](m),$$

whereas a^* was obtained via backtracking. Show that a^* can also be computed without backtracking using the matching function $\Delta_{\mathrm{DTW}}[\mathrm{Rev}(X), \mathrm{Rev}(Y)]$.
[**Hint:** Study the relation between optimal paths that align X with subsequences of Y and optimal paths that align $\mathrm{Rev}(X)$ with subsequences of $\mathrm{Rev}(Y)$.]

Solution to Exercise 7.8. Let $P = (p_1, \ldots, p_L)$ be a path with $p_\ell = (n_\ell, m_\ell) \in [1 : N] \times [1 : M]$, $\ell \in [1 : L]$, and $n_1 = 1$ and $n_L = N$. Define $Q = (q_1, \ldots, q_L)$ by setting

$$q_\ell := (q_\ell^1, q_\ell^2) := (N - n_{L-\ell+1} + 1, M - m_{L-\ell+1} + 1)$$

for $\ell \in [1 : L]$. Then, Q defines a path with $q_1 = (1, M - m_L + 1)$ and $q_L = (N, M - m_1 + 1)$. Next, we show that the total cost $c_P(X,Y)$ of the path P (see (3.20)) coincides with the total cost $c_Q(\mathrm{Rev}(X), \mathrm{Rev}(Y))$:

$$\begin{aligned} c_Q(\mathrm{Rev}(X),\mathrm{Rev}(Y)) &= \sum_{\ell=1}^{L} c(x_{N-q_\ell^1+1}, y_{M-q_\ell^2+1}) \\ &= \sum_{\ell=1}^{L} c(x_{N-(N-n_{L-\ell+1}+1)+1}, y_{M-(M-m_{L-\ell+1}+1)+1}) \\ &= \sum_{\ell=1}^{L} c(x_{n_{L-\ell+1}}, y_{m_{L-\ell+1}}) \\ &= \sum_{\ell=1}^{L} c(x_{n_\ell}, y_{m_\ell}) = c_P(X,Y). \end{aligned}$$

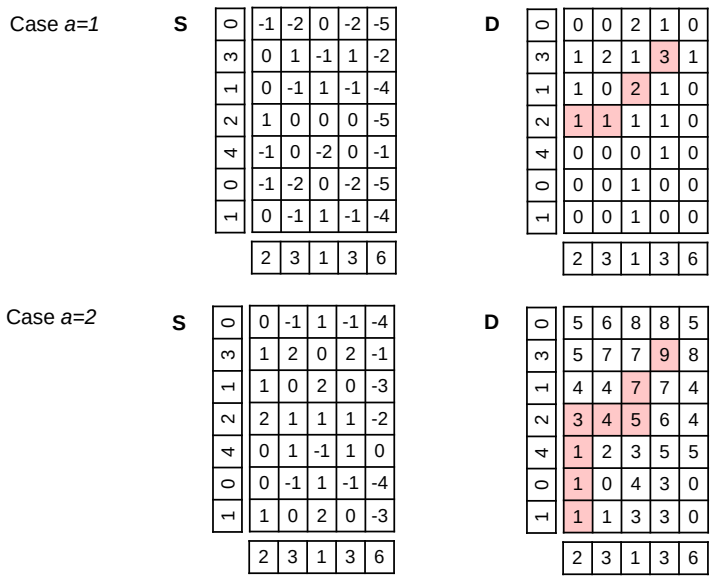
From this, it follows that P is an optimal warping path for the sequence X and the subsequence $Y(m_1 : m_L)$ of Y if and only if Q is an optimal warping path for the sequence $\mathrm{Rev}(X)$ and the subsquence $\mathrm{Rev}(Y)(M-m_L+1 : M-m_1+1)$ of $\mathrm{Rev}(Y)$. In particular, let $P = (p_1,\ldots,p_L)$ be the optimal warping path for the subsequence $Y(a^* : b^*)$, i.e., $p_1 = (1,a^*)$ and $p_L = (N,b^*)$. Then, $Q = (q_1,\ldots,q_L)$ is an optimal warping path for the subsequence $\mathrm{Rev}(Y)(M-b^*+1 : M-a^*+1)$. Therefore, minimizing over the matching funtion $\Delta_{\mathrm{DTW}}[\mathrm{Rev}(X),\mathrm{Rev}(Y)]$ yields the index $M-a^*+1$. In other words, a^* can be obtained via

$$a^* = M - \Big(\mathrm{argmin}_{m\in[1:M]}\, \Delta_{\mathrm{DTW}}[\mathrm{Rev}(X),\mathrm{Rev}(Y)](m)\Big) + 1.$$

Exercise 7.9. Let $\mathcal{F} = \mathbb{R}$ be a feature space and $s := s^a : \mathcal{F}\times\mathcal{F} \to \mathbb{R}$ a similarity measure defined by $s^a(x,y) := a - |x-y|$ for a constant $a \in \mathbb{R}$ and $x,y \in \mathbb{R}$ (see also Exercise 4.1). Given the sequences $X = (x_1,\ldots,x_N) = (1,0,4,2,1,3,0)$ of length $N = 7$ and $Y = (y_1,\ldots,y_M) = (2,3,1,3,6)$ of length $M = 5$, compute the optimal local alignment (best matching subsequences) of X and Y using the procedure described in Section 7.3.2. To this end, compute the similarity matrix $\mathbf{S}$ (see (7.31)) using $s = s^1$ (i.e., $a = 1$), the accumluated score matrix $\mathbf{D}$ (see (7.33)), the score-maximizing path P^* (see (7.32)), and the two induced segments $\pi_1(P^*)$ and $\pi_2(P^*)$ (see also Figure 7.20).

Then, in the same fashion, compute the optimal local alignment using the similarity measure $s = s^2$ (i.e., $a = 2$). What do you expect when further increasing the number a? Why is it problematic when all entries of $\mathbf{S}$ are positive?

Solution to Exercise 7.9. The following figure shows the similarity matrix $\mathbf{S}$ as well the accumluated cost matrix $\mathbf{D}$ for the case $a = 1$ as well as the case $a = 2$:



In the case $a = 1$, one obtains $P^* = ((4,1),(4,2),(5,3),(6,4))$, thus $\pi_1(P^*) = [4:6]$ and $\pi_2(P^*) = [1:4]$. Therefore, the best matching subsequences are $X(4:6)$ and $Y(1:4)$.

When using the parameter $a = 2$, one obtains $P^* = ((1,1),(2,1),(3,1),(4,1),(4,2),(4,3),(5,3),(6,4))$, thus $\pi_1(P^*) = [1:6]$ and $\pi_2(P^*) = [1:4]$. Therefore, the best matching subsequences are $X(1:6)$ and $Y(1:4)$. Increasing the parameter a makes the entries in $\mathbf{S}$ larger. As a result, the two best matching subsequences generally become longer. In the case that $\mathbf{S}$ has only positive entries, the best matching subsequences are the entire sequences X and Y (since in this case the accumluated score is optimized). Therefore, when being interested in capturing local similarities, it is important to find a good balance between positive and negative values in $\mathbf{S}$. As discussed in Section 7.3.2, only the cells that may express relevant similarity relations should have a positive score, whereas all other cells should have a negative score.

Exercise 7.10. Let $X = (x_1, x_2, \ldots, x_N)$ and $Y = (y_1, y_2, \ldots, y_M)$ be two sequences over the feature space $\mathcal{F}$. A **partial match** of length $L \in \mathbb{N}_0$ between X and Y is defined to be a sequence $P = ((n_1, m_1), \ldots, (n_L, m_L))$ of cells $(n_\ell, m_\ell) \in [1:N] \times [1:M]$, $\ell \in [1:L]$, which is strictly monotonically increasing:

$$n_1 < n_2 < \ldots < n_L \quad \text{and} \quad m_1 < m_2 < \ldots < m_L.$$

Given a similarity measure $s : \mathcal{F} \times \mathcal{F} \to \mathbb{R}$, define the similarity matrix $\mathbf{S}$ by $\mathbf{S}(n,m) := s(x_n, y_m)$ as in (7.31). Then, the total score $\sigma(P)$ of a partial match P is specified by

$$\sigma(P) := \sum_{\ell=1}^{L} \mathbf{S}(n_\ell, m_\ell).$$

Describe an algorithm based on dynamic programming as in Table 3.2 to compute an optimal (i.e., score-maximizing) partial match.

Solution to Exercise 7.10. The following table shows a specification of the algorithm based on dynamic programming for computing an optimal partial match:

Algorithm: PARTIAL MATCHING

Input: Similarity matrix $\mathbf{S}$ of size $N \times M$
Output: Accumulated score matrix $\mathbf{D}$
Optimal partial match P^*

Procedure: Initialize an $((N+1)\times(M+1))$ matrix $\mathbf{D}$ by $D(0,0) = D(n,0) = D(0,m) = 0$ for $n \in [1:N]$ and $m \in [1:M]$. Then compute in a nested loop for $n = 1,\ldots,N$ and $m = 1,\ldots,M$:

$$\mathbf{D}(n,m) = \max\{\mathbf{D}(n,m-1), \mathbf{D}(n-1,m), \mathbf{S}(n,m) + \mathbf{D}(n-1,m-1)\}.$$

Set $n = N$, $m = M$, $\ell = 0$. Then repeat the following steps while $(n > 0)$ and $(m > 0)$:

If $D(n,m) = D(n,m-1)$	then $m := m-1$
else if $D(n,m) = D(n-1,m)$	then $n := n-1$
else if $D(n,m) = \mathbf{S}(n,m) + D(n-1,m-1)$	then $\ell := \ell+1$, $q(\ell) := (n,m)$, and $n := n-1$, $m := m-1$

Set $L = \ell$ and return $P^* = (q_L, q_{L-1}, \ldots, q_1)$ as well as $\sigma(P^*) = \mathbf{D}(N,M)$.

Note that in the nested loop, the score value $\mathbf{S}(n,m)$ is only added in the case of a diagonal step size—as opposed to the DTW algorithm (see Table 3.2), where the cost value $\mathbf{C}(n,m)$ is added in the case of all three step sizes. Furthermore, note that the optimal partial match yielding the accumulated score $\mathbf{D}(n,m)$ does *not* necessarily end in the cell (n,m)—in contrast with the DTW algorithm, where the optimal warping path yielding $\mathbf{D}(n,m)$ always ends with the cell (n,m).

Exercise 7.11. Show that the definitions of the precision $\mathrm{P}_{\mathcal{Q}}(r)$ and recall $\mathrm{R}_{\mathcal{Q}}(r)$ at rank $r \in [1:K]$ in (7.45) and (7.46) agree with the definitions in (4.47) and (4.48), respectively. To this end, depending on r, define a suitable set $\mathcal{I}_+^{\mathrm{Est}}$.

Solution to Exercise 7.11. Recall from Section 7.3.3 that the set of items of our retrieval scenario is $\mathcal{I} = [1:K]$, which represents a collection that consists of K database documents. Furthermore, the set of relevant or positive items (reference annotations) is denoted by $\mathcal{I}_+^{\mathrm{Ref}} = \mathcal{I}_{\mathcal{Q}}$ (see (7.43)). For a fixed rank $r \in [1:K]$, let

$$\mathcal{I}_+^{\mathrm{Est}} := \big\{\rho_{\mathcal{Q}}(1), \rho_{\mathcal{Q}}(2), \ldots, \rho_{\mathcal{Q}}(r)\big\}$$

be the set of the top r items of the ranked list. We consider this set to be the set of items estimated as positive (see Section 4.5.1). Now, by definition (7.44) of the relevance function $\chi_{\mathcal{Q}}$, the sum $\sum_{k=1}^{r} \chi_{\mathcal{Q}}(k)$ corresponds to the number of relevant items among the top r items. Thus, we obtain

$$|\mathcal{I}_+^{\mathrm{Est}} \cap \mathcal{I}_+^{\mathrm{Ref}}| = \sum_{k=1}^{r} \chi_{\mathcal{Q}}(k).$$

From this and using $|\mathcal{I}_+^{\mathrm{Est}}| = r$ and $|\mathcal{I}_+^{\mathrm{Ref}}| = |\mathcal{I}_\mathcal{Q}|$, we obtain

$$\mathrm{P}_\mathcal{Q}(r) = \frac{1}{r}\sum_{k=1}^{r}\chi_\mathcal{Q}(k) = \frac{|\mathcal{I}_+^{\mathrm{Est}} \cap \mathcal{I}_+^{\mathrm{Ref}}|}{|\mathcal{I}_+^{\mathrm{Est}}|},$$
$$\mathrm{R}_\mathcal{Q}(r) = \frac{1}{|\mathcal{I}_\mathcal{Q}|}\sum_{k=1}^{r}\chi_\mathcal{Q}(k) = \frac{|\mathcal{I}_+^{\mathrm{Est}} \cap \mathcal{I}_+^{\mathrm{Ref}}|}{|\mathcal{I}_+^{\mathrm{Ref}}|},$$

which shows that (7.45) coincides with (4.47) and (7.46) with (4.48).

Exercise 7.12. Let us consider a database $\{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_K\}$ consisting of $K = 8$ documents. Given a query document $\mathcal{Q}$, assume that we have a similarity measure that yields the following values $\gamma(\mathcal{Q}, \mathcal{D}_k) \in \mathbb{R}$ for each $k \in [1:K]$:

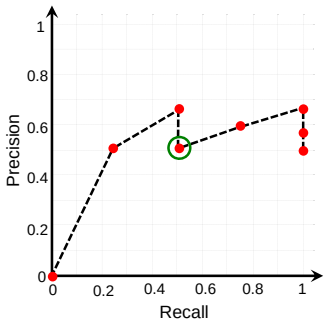
k	1	2	3	4	5	6	7	8
$\gamma(\mathcal{Q}, \mathcal{D}_k)$	0.7	2.6	3.6	3.5	3.2	3.7	1.5	3.1

Furthermore, let $\mathcal{I}_\mathcal{Q} = \{2,3,4,8\}$ be the set of the relevant items (see (7.43)). Calculate the precision $\mathrm{P}_\mathcal{Q}(r)$ and recall $\mathrm{R}_\mathcal{Q}(r)$ at rank $r \in [1:K]$. Furthermore, draw the corresponding precision–recall curve (as in Figure 7.21c). Finally, determine the break-even point, the maximal F-measure $\mathrm{F}_\mathcal{Q}^{\max}$ (see (7.47)), as well as the average precision $\overline{\mathrm{P}}_\mathcal{Q}$ (see (7.48)).

Solution to Exercise 7.12. The following figure shows the precision $\mathrm{P}_\mathcal{Q}(r)$ and recall $\mathrm{R}_\mathcal{Q}(r)$ at rank $r \in [1:K]$ (rounded to two decimal places) as well as the resulting PR curve:

ID	Score	Rank
1	0.7	8
2	2.6	6
3	3.6	2
4	3.5	3
5	3.2	4
6	3.7	1
7	1.5	7
8	3.1	5

Rank	ID	Rel.	P(r)	R(r)	F(r)
1	6	-	**0**	**0**	**0**
2	3	+	**0.50**	**0.25**	**0.33**
3	4	+	**0.67**	**0.50**	**0.57**
4	5	-	**0.50**	**0.50**	**0.50**
5	8	+	**0.60**	**0.75**	**0.67**
6	2	+	**0.67**	**1.00**	**0.80**
7	7	-	**0.57**	**1.00**	**0.73**
8	1	-	**0.50**	**1.00**	**0.67**



The break-even point is $\mathrm{P}_\mathcal{Q}(4) = \mathrm{R}_\mathcal{Q}(4) = 0.5$. Furthermore, the maximal F-measure is $\mathrm{F}_\mathcal{Q}^{\max} = 0.8$ and the average precision is

$$\overline{\mathrm{P}}_\mathcal{Q} = \frac{1}{4}(1/2 + 2/3 + 3/5 + 2/3) = \frac{73}{120} \approx 0.6083.$$

Exercise 7.13. Let us consider a PR curve $\{(\mathrm{P}_\mathcal{Q}(r), \mathrm{R}_\mathcal{Q}(r)) \mid r \in [1:K]\}$ for a ranked retrieval result over K database documents. Recall that the break-even point of the PR curve is the positive value where the precision equals the recall. Show that the

break-even point exists if and only if there is at least one relevant document among the top $|\mathcal{I}_\mathcal{Q}|$ items of the ranked list. Furthermore, show that in this case $\mathrm{P}_\mathcal{Q}(r) = \mathrm{R}_\mathcal{Q}(r)$ if and only if $r = |\mathcal{I}_\mathcal{Q}|$.

Solution to Exercise 7.13. Assume that a break-even point exists. Then, $\mathrm{P}_\mathcal{Q}(r) = \mathrm{R}_\mathcal{Q}(r) > 0$ for some $r \in [1:K]$. Using definitions (7.45) and (7.46), we obtain

$$\frac{1}{r}\sum_{k=1}^{r}\chi_\mathcal{Q}(k) = \mathrm{P}_\mathcal{Q}(r) = \mathrm{R}_\mathcal{Q}(r) = \frac{1}{|\mathcal{I}_\mathcal{Q}|}\sum_{k=1}^{r}\chi_\mathcal{Q}(k) > 0.$$

In particular, this implies that $\sum_{k=1}^{r}\chi_\mathcal{Q}(k) > 0$ (i.e., there is at least one relevant item among the top r items of the ranked list) as well as $r = |\mathcal{I}_\mathcal{Q}|$ (since $\sum_{k=1}^{r}\chi_\mathcal{Q}(k) > 0$). t Now, assume that there is at least one relevant document among the top $|\mathcal{I}_\mathcal{Q}|$ items of the ranked list. Then, $\sum_{k=1}^{|\mathcal{I}_\mathcal{Q}|}\chi_\mathcal{Q}(k) > 0$. Furthermore, using definitions (7.45) and (7.46), one obtains $\mathrm{P}_\mathcal{Q}(r) = \mathrm{R}_\mathcal{Q}(r) > 0$ for $r = |\mathcal{I}_\mathcal{Q}|$, i.e., there exists a break-even point.

Exercise 7.14. Show that the maximal F-measure of a PR curve is at least as large as the break-even point (if it exists; see Exercise 7.13). Give an example where the maximal F-measure and the break-even point do not coincide.

Solution to Exercise 7.14. Assume that the break-even point exists. Then, by Exercise 7.13, the break-even point assumed for $r = |\mathcal{I}_\mathcal{Q}|$. From $\mathrm{P}_\mathcal{Q}(r) = \mathrm{R}_\mathcal{Q}(r)$, it follows that $\mathrm{F}_\mathcal{Q}(r) = \mathrm{P}_\mathcal{Q}(r) = \mathrm{R}_\mathcal{Q}(r)$. Since $\mathrm{F}_\mathcal{Q}^{\max} \geq \mathrm{F}_\mathcal{Q}(r)$, this shows that the maximal F-measure is at least as large as the break-even point. In Exercise 7.12, we have already seen an example where the maximal F-measure and the break-even point do not coincide. As another example, consider the relevance function $\chi_\mathcal{Q} : [1:K] \to \{0,1\}$ for a ranked retrieval result for $K = 3$ defined by $\chi_\mathcal{Q}(1) = 1$, $\chi_\mathcal{Q}(2) = 0$, and $\chi_\mathcal{Q}(3) = 1$. Then, we obtain

r	$\mathrm{P}_\mathcal{Q}(r)$	$\mathrm{R}_\mathcal{Q}(r)$	$\mathrm{F}_\mathcal{Q}(r)$
1	1	1/2	2/3
2	1/2	1/2	1/2
3	2/3	1	4/5

From this we obtain a break-even point of 0.5 and a maximal F-measure of $\mathrm{F}_\mathcal{Q}^{\max} = 0.8$.

Exercise 7.15. Let us consider a database consisting of $K \in \mathbb{N}$ documents. Furthermore, let $\mathcal{Q}$ be a query document with $L := |\mathcal{I}_\mathcal{Q}| \in [1:K]$ relevant items. Assume that the relevant items are ranked by a retrieval system at the positions

$$r_1 < r_2 < \ldots < r_L,$$

where $r_\ell \in [1:K]$ for $\ell \in [1:L]$. (Recall from Section 7.3.3 that, the smaller the index r_ℓ, the higher the rank of the document.) Specify a formula for the average precision $\overline{\mathrm{P}}_\mathcal{Q}$ of this ranking (see (7.48)). Furthermore, assuming $K = 5$ and $L = 2$, calculate the average precision for all possible rankings.

Solution to Exercise 7.15. Recall from (7.44) that the relevance function $\chi_{\mathcal{Q}} : [1:K] \to \{0,1\}$ assumes the value $\chi_{\mathcal{Q}}(r) = 1$ for some $r \in [1:K]$ if and only if the document at rank r is relevant. Therefore, in our scenario, $\chi_{\mathcal{Q}}(r) = 1$ if and only if $r = r_\ell$ for some $\ell \in [1:L]$. Furthermore, note that one has the precision $\mathrm{P}_{\mathcal{Q}}(r_\ell) = \ell / r_\ell$ at rank r_ℓ for $\ell \in [1:L]$. From this and (7.48), the average precision computes as

$$\overline{\mathrm{P}}_{\mathcal{Q}} = \frac{1}{|\mathcal{I}_{\mathcal{Q}}|} \sum_{r=1}^{K} \mathrm{P}_{\mathcal{Q}}(r) \chi_{\mathcal{Q}}(r) = \frac{1}{L} \sum_{\ell=1}^{L} \mathrm{P}_{\mathcal{Q}}(r_\ell) = \frac{1}{L} \sum_{\ell=1}^{L} \frac{\ell}{r_\ell}.$$

There are $K!$ different rankings (corresponding to the number of permutations of $[1:K]$) for a set of K documents. However, as with the average precision, the ranking only depends on the ranking positions

$$r_1 < r_2 < \ldots < r_L$$

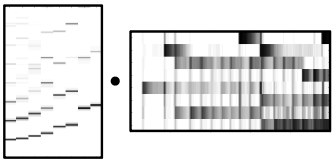
of the relevant documents. In general, there are $\binom{K}{L}$ possibilities for such ranking positions. In the case $K = 5$ and $L = 2$, the ranking positions are represented by pairs (r_1, r_2) with $r_1, r_2 \in [1:5]$ and $r_1 < r_2$. Note that there are $\binom{5}{2} = 10$ such pairs. The average precision for (r_1, r_2) is given by

$$\overline{\mathrm{P}}_{\mathcal{Q}} = \frac{1}{2} \left(\frac{1}{r_1} + \frac{2}{r_2} \right).$$

This yields the following values:

(r_1, r_2)	(1,2)	(1,3)	(1,4)	(1,5)	(2,3)	(2,4)	(2,5)	(3,4)	(3,5)	(4,5)
$\overline{\mathrm{P}}_{\mathcal{Q}}$	1.000	0.833	0.750	0.700	0.583	0.500	0.450	0.417	0.367	0.325

Chapter 8
Musically Informed Audio Decomposition



Exercise 8.1. The arithmetic mean $\mu(A)$ of a list $A = (a_1, a_2, \ldots, a_L)$ that consists of real numbers $a_\ell \in \mathbb{R}$, $\ell \in [1:L]$ is defined by $\mu(A) := \big(\sum_{\ell=1}^{L} a_\ell\big)/L$. Let $A = (2,3,190,2,3)$. Compute the mean $\mu(A)$ as well as the median $\mu_{1/2}(A)$ (see (8.4)). Explain why the HPS algorithm described in Section 8.1 employs median filtering and not mean filtering.

Solution to Exercise 8.1. One obtains $\mu(A) = 40$ and $\mu_{1/2}(A) = 3$. The median is robust to outliers, whereas the mean is heavily influenced by a small number of extreme values. As illustrated by Figure 8.4, percussive events can be regarded as outliers across time at for a given frequency parameter. Therefore, to remove the percussive events, median filtering in the horizontal (time) direction is more suitable than mean filtering. Similarly, harmonic events can be regarded as outliers across frequency at a given time frame, again justifying the usage of median filtering.

Exercise 8.2. Let

$$\mathcal{Y} = \begin{bmatrix} 1 & 1 & 46 & 2 \\ 3 & 1 & 50 & 1 \\ 60 & 68 & 70 & 67 \\ 2 & 1 & 65 & 1 \end{bmatrix}$$

be a spectrogram. Assuming a suitable zero-padding, compute $\tilde{\mathcal{Y}}^{\mathrm{h}}$ as in (8.6) using $L^{\mathrm{h}} = 3$ and $\tilde{\mathcal{Y}}^{\mathrm{p}}$ as in (8.7) using $L^{\mathrm{p}} = 3$. Furthermore, compute the binary mask $\mathcal{M}^{\mathrm{h}}$ as in (8.8) and $\mathcal{M}^{\mathrm{p}}$ as in (8.9). Finally, apply the masks to the matrix $\mathcal{Y}$ using pointwise multiplication to derive the two matrices $\mathcal{Y}^{\mathrm{h}}$ as in (8.12) and $\mathcal{Y}^{\mathrm{p}}$ as in (8.13).

Solution to Exercise 8.2. The filtered matrices are as follows:

$$\tilde{\mathcal{Y}}^{\mathrm{h}} = \begin{bmatrix} 1 & 1 & 2 & 2 \\ 1 & 3 & 1 & 1 \\ 60 & 68 & 68 & 67 \\ 1 & 2 & 1 & 1 \end{bmatrix} \quad \text{and} \quad \tilde{\mathcal{Y}}^{\mathrm{p}} = \begin{bmatrix} 1 & 1 & 46 & 1 \\ 3 & 1 & 50 & 2 \\ 3 & 1 & 65 & 1 \\ 2 & 1 & 65 & 1 \end{bmatrix}.$$

From this one obtains the following binary masks:

$$\mathcal{M}^{\mathrm{h}} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathcal{M}^{\mathrm{p}} = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{bmatrix}$$

Pointwise multiplication yields the following masked spectrograms:

$$\mathcal{Y}^{\mathrm{h}} = \begin{bmatrix} 1 & 1 & 0 & 2 \\ 0 & 3 & 0 & 0 \\ 60 & 68 & 68 & 67 \\ 0 & 2 & 0 & 1 \end{bmatrix} \quad \text{and} \quad \mathcal{Y}^{\mathrm{p}} = \begin{bmatrix} 0 & 0 & 46 & 0 \\ 3 & 0 & 50 & 2 \\ 0 & 0 & 0 & 0 \\ 2 & 0 & 65 & 0 \end{bmatrix}$$

Exercise 8.3. Let F_{s} (given in Hz) be the sampling rate of a given signal x. Furthermore, let $N \in \mathbb{N}$ be the window length and $H \in \mathbb{N}$ the hop size of a discrete STFT. The filter lengths $L^{\mathrm{h}}, L^{\mathrm{p}} \in \mathbb{N}$ of the median filters used in the HPS approach are specified in terms of frame and frequency indices of the underlying STFT. In practice, it may be more convenient if a user can specify the filter length L^{h} in terms of seconds and L^{p} in terms of Hertz. Derive a formula that converts a time duration $\Delta_t \in \mathbb{R}$ given in seconds to a minimum filter length $L^{\mathrm{h}}(\Delta_t) \in \mathbb{N}$ given in frame indices covering this duration. Similarly, derive a formula that converts a frequency range $\Delta_\omega \in \mathbb{R}$ given in Hertz to a minimum filter length $L^{\mathrm{p}}(\Delta_\omega) \in \mathbb{N}$ given in frequency indices covering this range. Finally, assuming $F_{\mathrm{s}} = 22050$ Hz, $N = 1024$, and $H = 256$, determine $L^{\mathrm{h}}(\Delta_t)$ for $\Delta_t = 0.5$ sec and $L^{\mathrm{p}}(\Delta_\omega)$ for $\Delta_\omega = 600$ Hz.

Solution to Exercise 8.3. There is a frame every H/F_{s} seconds (see (2.27)). Therefore, we obtain

$$L^{\mathrm{h}}(\Delta_t) = \left\lceil \Delta_t \cdot \frac{F_{\mathrm{s}}}{H} \right\rceil .$$

Furthermore, each frequency parameter corresponds to F_{s}/N Hertz (see (2.28)). Therefore, we obtain

$$L^{\mathrm{p}}(\Delta_\omega) = \left\lceil \Delta_\omega \cdot \frac{N}{F_{\mathrm{s}}} \right\rceil .$$

These formulas yield the values $L^{\mathrm{h}}(0.5) = 44$ and $L^{\mathrm{p}}(600) = 28$.

Exercise 8.4. Show that one obtains a partition of unity (see (8.22)) when using the discrete window function $w : \mathbb{Z} \to \mathbb{R}$ defined by

$$w(r) := \begin{cases} \sin(\pi r/N)^2 & \text{if } r \in [0 : N-1], \\ 0 & \text{otherwise} \end{cases}$$

(see (8.23)) and the hop size $H = N/2$ (assuming that N is even). What happens if the hop size $H = N/4$ (assuming that N is divisible by four) is used instead? Give a proof of your claim.

Solution to Exercise 8.4. As preparation for the proof, let $\mathbf{s}_N^2 : \mathbb{Z} \to \mathbb{R}$ be defined by $\mathbf{s}_N^2(r) := \sin(\pi r/N)^2$ and $\mathbf{c}_N^2 : \mathbb{Z} \to \mathbb{R}$ by $\mathbf{c}_N^2(r) := \cos(\pi r/N)^2$ for $r \in \mathbb{Z}$. These two functions are periodic with period N (since we are considering the squared versions of sin and cos) and, by a trigonemtric identity, one obtains

$$\mathbf{s}_N^2(r) + \mathbf{c}_N^2(r) = 1$$

for all $r \in \mathbb{Z}$. Note that the window w from (8.23) coincides with $\mathbf{s}_N^2$ on the interval $[0:N-1]$. It follows that

$$\sum_{n\in\mathbb{Z}} w(r-nN) = \mathbf{s}_N^2(r).$$

Furthermore, from the relation $\mathbf{s}_N^2(r-N/2) = \mathbf{c}_N^2(r)$, it follows that

$$\sum_{n\in\mathbb{Z}} w(r-nN-N/2) = \mathbf{c}_N^2(r).$$

After these preparations, we now show that w and its translates define a partition of unity when using the hop size $H = N/2$ (see (8.22)). To this end, in the following computation, we split up the sum over $n \in \mathbb{Z}$ into a sum over even integers and a sum over odd integers:

$$\begin{aligned}
\sum_{n\in\mathbb{Z}} w(r-nH) &= \sum_{n\in\mathbb{Z}} w(r-nN/2) \\
&= \sum_{n\in\mathbb{Z}} w(r-(2n)N/2) + \sum_{n\in\mathbb{Z}} w(r-(2n+1)N/2) \\
&= \sum_{n\in\mathbb{Z}} w(r-nN) + \sum_{n\in\mathbb{Z}} w(r-nN-N/2) \\
&= \mathbf{s}_N^2(r) + \mathbf{c}_N^2(r) \\
&= 1.
\end{aligned}$$

As for the hop size $H = N/4$, we again split up the summation over even n and odd n. Summation over all even n gives the same result as in the previous case (using $H = N/2$ and summing over all n). Therefore, this sum is one. Similarly, summing over all odd n yields the same function as in the even case, up to a shift of $H = N/4$. Again, the sum is one. Therefore, it follows that

$$\sum_{n\in\mathbb{Z}} w(r-nN/4) = \sum_{n\in\mathbb{Z}} w(r-nN/2) + \sum_{n\in\mathbb{Z}} w(r-nN/2-N/4) = 2$$

for $r \in \mathbb{Z}$.

Exercise 8.5. One problem in harmonic–percussive separation (HPS) is that a sound may contain noise-like events (e.g., applause, distorted guitar) that are neither of harmonic nor of percussive nature. In this exercise, we study an extension to HPS by considering a third residual component which captures the sounds that lie "between" a clearly harmonic and a clearly percussive component. To this end, we introduce an additional parameter $\beta \in \mathbb{R}$ with $\beta \geq 1$ called the **separation factor**. Generalizing (8.8) and (8.9), we define the binary masks $\mathcal{M}^{\mathrm{h}}$, $\mathcal{M}^{\mathrm{p}}$, and $\mathcal{M}^{\mathrm{r}}$ for the clearly harmonic, the clearly percussive, and the residual components by setting

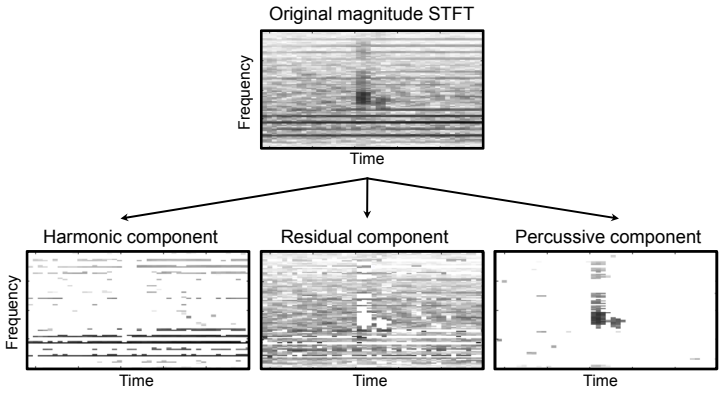
$$\mathcal{M}^{\mathrm{h}}(n,k) := \begin{cases} 1 & \text{if } \tilde{\mathcal{Y}}^{\mathrm{h}}(n,k) \geq \beta \cdot \tilde{\mathcal{Y}}^{\mathrm{p}}(n,k), \\ 0 & \text{otherwise,} \end{cases}$$
$$\mathcal{M}^{\mathrm{p}}(n,k) := \begin{cases} 1 & \text{if } \tilde{\mathcal{Y}}^{\mathrm{p}}(n,k) > \beta \cdot \tilde{\mathcal{Y}}^{\mathrm{h}}(n,k), \\ 0 & \text{otherwise,} \end{cases}$$
$$\mathcal{M}^{\mathrm{r}}(n,k) := 1 - \big(\mathcal{M}^{\mathrm{h}}(n,k) + \mathcal{M}^{\mathrm{p}}(n,k)\big).$$

Using these masks, derive a signal decomposition $x = x^{\mathrm{h}} + x^{\mathrm{p}} + x^{\mathrm{r}}$. Furthermore, discuss the role of the parameter β. How do the components change when successively increasing β?

Solution to Exercise 8.5. As in (8.14) and (8.15), one defines

$$\mathcal{X}^{\mathrm{h}}(n,k) := \mathcal{M}^{\mathrm{h}}(n,k) \cdot \mathcal{X}(n,k),$$
$$\mathcal{X}^{\mathrm{p}}(n,k) := \mathcal{M}^{\mathrm{h}}(n,k) \cdot \mathcal{X}(n,k),$$
$$\mathcal{X}^{\mathrm{r}}(n,k) := \mathcal{M}^{\mathrm{r}}(n,k) \cdot \mathcal{X}(n,k).$$

The following figure shows a typical example for the resulting decomposition of a given STFT:



From this, one can derive x^{h}, x^{p}, and x^{r} by applying an inverse STFT (using an overlap–add technique as described in Section 8.1.2.2). The separation factor β can be used to adjust the decomposition. The case $\beta = 1$ translates to the original HPS decomposition. By increasing β, less time–frequency bins are assigned for the reconstruction of the components x^{h} and x^{p}, whereas more time–frequency bins are used for the reconstruction of the residual component x^{r}. Intuitively, the larger the parameter β, the clearer becomes the harmonic and percussive nature of the components x^{h} and x^{p}. For very large β, the residual signal x^{r} tends to contain the entire signal x. For further details, see

J. DRIEDGER, M. MÜLLER, AND S. DISCH, *Extending harmonic–percussive separation of audio signals*, in Proceedings of the 15th International Conference on Music Information Retrieval (ISMIR), Taipei, Taiwan, 2014, pp. 611–616.

Exercise 8.6. Derive the formula (8.44) for the instantaneous frequency $F_{\text{coef}}^{\text{IF}}(k,n)$ and the formula (8.45) for the bin offset $\kappa(k,n)$.

Solution to Exercise 8.6. From (8.40), (8.41), and (8.42), we have:

$$\begin{aligned}
\omega &= F_{\text{coef}}(k) = \frac{k \cdot F_{\text{s}}}{N}, \\
t_1 &= T_{\text{coef}}(n-1) = \frac{(n-1)\cdot H}{F_{\text{s}}} \quad \text{and} \quad t_2 = T_{\text{coef}}(n) = \frac{n \cdot H}{F_{\text{s}}}, \\
\varphi_1 &= \varphi(n-1,k) \quad \text{and} \quad \varphi_2 = \varphi(n,k).
\end{aligned}$$

From this, (8.32), and (8.33), we obtain the following equations:

$$\begin{aligned}
\Delta t &= t_2 - t_1 = T_{\text{coef}}(n) - T_{\text{coef}}(n-1) = \frac{H}{F_{\text{s}}}, \\
\varphi^{\text{Pred}} &= \varphi_1 + \omega \cdot \Delta t = \varphi(n-1,k) + \frac{k \cdot F_{\text{s}}}{N} \cdot \frac{H}{F_{\text{s}}} = \varphi(n-1,k) + \frac{k \cdot H}{N}, \\
\varphi^{\text{Err}} &= \Psi(\varphi_2 - \varphi^{\text{Pred}}) = \Psi\left(\varphi(n,k) - \varphi(n-1,k) - \frac{k \cdot H}{N}\right).
\end{aligned}$$

Using (8.34), we obtain:

$$F_{\text{coef}}^{\text{IF}}(k,n) = \omega + \frac{\varphi^{\text{Err}}}{\Delta t} = \frac{k \cdot F_{\text{s}}}{N} + \frac{\varphi^{\text{Err}} \cdot F_{\text{s}}}{H} = \left(k + \frac{N}{H} \cdot \varphi^{\text{Err}}\right) \cdot \frac{F_{\text{s}}}{N} = (k + \kappa(k,n)) \cdot \frac{F_{\text{s}}}{N}$$

with

$$\kappa(k,n) = \frac{N}{H} \cdot \Psi\left(\varphi(n,k) - \varphi(n-1,k) - \frac{k \cdot H}{N}\right).$$

Exercise 8.7. We have seen in Section 8.2.1 that the quality of the estimated instantaneous frequency depends on the length $\Delta t = t_2 - t_1 = H/F_{\text{s}}$. Therefore, it is beneficial to use a small hop size H. On the downside, using a small hop size increases the computational cost for calculating the discrete STFT. An alternative approach for obtaining good instantaneous frequency estimates is to keep the original hop size, but compute the STFT twice—the second time at a lag of just one sample. Discuss the benefits of this alternative approach over the strategy of simply reducing the hop size.

Solution to Exercise 8.7. In the alternative approach, one can keep the original hop size H, but obtain instantaneous frequency corrections based on the smallest possible hop size (one frame). The computational cost only doubles. In constrast, reducing the hop size to, e.g., $H/4$, the computational cost would increase by a factor of four.

Exercise 8.8. Defining $\text{Bin}(\omega) := \lfloor 12 \cdot \log_2(\omega/440) + 69.5 \rfloor$ for $\omega \in \mathbb{R}$ as in (8.47), show that $\omega \in [F_{\text{pitch}}(p-0.5), F_{\text{pitch}}(p+0.5))$ if and only if $\text{Bin}(\omega) = p$ for $p \in \mathbb{Z}$.

Solution to Exercise 8.8.

$$\begin{aligned}
& \omega \in [F_{\text{pitch}}(p-0.5), F_{\text{pitch}}(p+0.5)) \\
\iff \quad & 2^{(p-69.5)/12} \cdot 440 \leq \omega < 2^{(p-68.5)/12} \cdot 440 \\
\iff \quad & p - 69.5 \leq 12 \cdot \log_2\left(\frac{\omega}{440}\right) < p - 68.5 \\
\iff \quad & p \leq 12 \cdot \log_2\left(\frac{\omega}{440}\right) + 69.5 < p + 1 \\
\iff \quad & p = \left\lfloor 12 \cdot \log_2\left(\frac{\omega}{440}\right) + 69.5 \right\rfloor \\
\iff \quad & \text{Bin}(\omega) = p
\end{aligned}$$

Exercise 8.9. Let ω be a frequency and $h \cdot \omega$ its h^{th} harmonic for some $h \in \mathbb{N}$. Considering the bin mapping function from (8.49), determine the relation between $\text{Bin}(\omega)$ and $\text{Bin}(h \cdot \omega)$. This relation explains the formula in (8.55) for the harmonic summation in the log-frequency domain.

Solution to Exercise 8.9. From (8.49), one obtains the following:

$$\begin{aligned}
\text{Bin}(h \cdot \omega) &= \left\lfloor \frac{1200}{R} \cdot \log_2\left(\frac{h \cdot \omega}{\omega_{\text{ref}}}\right) + 1.5 \right\rfloor \\
&= \left\lfloor \frac{1200}{R} \cdot \log_2(h) + \frac{1200}{R} \cdot \log_2\left(\frac{\omega}{\omega_{\text{ref}}}\right) + 1.5 \right\rfloor \\
&= \left\lfloor \frac{1200}{R} \cdot \log_2(h) \right\rfloor + \left\lfloor \frac{1200}{R} \cdot \log_2\left(\frac{\omega}{\omega_{\text{ref}}}\right) + 1.5 \right\rfloor + \delta \\
&= \left\lfloor \frac{1200}{R} \cdot \log_2(h) \right\rfloor + \text{Bin}(\omega) + \delta,
\end{aligned}$$

where $\delta \in \{0, 1\}$ is introduced to account for possible rounding inaccuracies.

Exercise 8.10. Let $\mathcal{Y}$ be a magnitude spectrogram with coefficients $\mathcal{Y}(n,k)$ for $n \in \mathbb{Z}$ and $k \in [0:K]$. Furthermore, for a given reference frequency ω_{ref} and a resolution R, let $\mathcal{Y}_{\text{LF}}$ be the log-frequency magnitude spectrogram as defined in (8.51) with coefficients $\mathcal{Y}_{\text{LF}}(n,b)$ for $n \in \mathbb{Z}$ and $b \in [1:B]$. Given a frequency trajectory $\eta : \mathbb{Z} \to [0:K]$ for $\mathcal{Y}$, describe how one can derive a corresponding trajectory $\eta_{\text{LF}} : \mathbb{Z} \to [1:B]$ for $\mathcal{Y}_{\text{LF}}$. Which problems may occur in this calculation?

Furthemore, let η^h and η^h_{LF} be the frequency trajectories of the first $H \in \mathbb{N}$ harmonics, $h \in [1:H]$. Note that $\eta^1 = \eta$ and $\eta^1_{\text{LF}} = \eta_{\text{LF}}$. Describe the mathematical relations between these trajectories. Thinking of practical computations and real-world musical sounds, discuss some problems that may introduce inaccuracies in these relations.

Solution to Exercise 8.10. Let $F_{\text{coef}}(k)$ be the center frequency for frequency index $k \in [0:K]$ (see (8.30)). Using the bin index function Bin from (8.49), we obtain η_{LF} from η via

$$\eta_{\text{LF}}(n) = \text{Bin}\big(F_{\text{coef}}(\eta(n))\big).$$

One problem in this calculation is that the resulting bin index may lie outside the range $[1:B]$, e.g., in the case that $F_{\text{coef}}(\eta(n)) < \omega_{\text{ref}}$ for some $n \in \mathbb{Z}$. This leaves $\eta_{\text{LF}}(n)$ undefined.

For the harmonics, we basically obtain

$$\eta^h(n) = \eta(n) \cdot h$$

as long as $\eta(n) \cdot h \leq K$. In the log-frequency domain, we obtain from (8.55) the relations

$$\eta^h_{\text{LF}}(n) = \eta_{\text{LF}}(n) + \left\lfloor \frac{1200}{R} \log_2(h) \right\rfloor.$$

Note that in these relations there may be increased inaccuracies for higher harmonics due to an accumulation of quantization errors, which result from the frequency grid introduced by the STFT and the binning. Furthermore, inharmonicities (i.e., deviations of partials from the closest ideal harmonics, see Section 1.3.2), may introduce further inaccuracies in these relations. For higher harmonics, such inaccuracies may become quite substantial for instruments such as the piano (see Section 1.3.4).

Exercise 8.11. The goal of this exercise is to develop an efficient algorithm for computing a frequency trajectory with temporal continuity constraints (see Section 8.2.3.1). Given a salience representation $\mathcal{Z} \in \mathbb{R}^{N \times B}_{\geq 0}$ and a transition likelihood matrix $\mathbf{T} \in \mathbb{R}^{B \times B}_{\geq 0}$, let $\sigma(\eta)$ be the total score for a given trajectory $\eta : [1:N] \to [1:B]$ as defined in (8.60). Specify an algorithm based on dynamic programming and backtracking (similar to the Viterbi algorithm in Table 5.2) for determining the score-maximizing trajectory η^{DP} (see (8.61)).

Solution to Exercise 8.11.

Algorithm: FREQUENCY TRAJECTORY WITH TEMPORAL CONTINUITY CONSTRAINT

Input: Salience representation $\mathcal{Z} \in \mathbb{R}^{N \times B}_{\geq 0}$
Transition likelihood matrix $\mathbf{T} \in \mathbb{R}^{B \times B}_{\geq 0}$

Output: Score-maximizing trajectory $\eta^{\text{DP}} : [1:N] \to [1:B]$

Procedure: Initialize the $(B \times N)$ matrix $\mathbf{D}$ by $\mathbf{D}(b,1) = \mathcal{Z}(1,b)$ for $b \in [1:B]$. Then compute in a nested loop for $n = 2, \ldots, N$ and $b = 1, \ldots, B$:

$$\begin{aligned} \mathbf{D}(b,n) &= \max_{c \in [1:B]} \big(\mathbf{T}(c,b) \cdot \mathbf{D}(c,n-1)\big) \cdot \mathcal{Z}(n,b) \\ \mathbf{E}(b,n-1) &= \operatorname{argmax}_{c \in [1:B]} \big(\mathbf{T}(c,b) \cdot \mathbf{D}(c,n-1)\big) \end{aligned}$$

Set $b_N = \operatorname{argmax}_{c \in [1:B]} \mathbf{D}(j,N)$ and compute for decreasing $n = N-1, \ldots, 1$ the maximizing indices

$$b_n = \mathbf{E}(b_{n+1}, n).$$

The optimal frequency trajectory η^{DP} is defined by $\eta^{\text{DP}}(n) = b_n$ for $n \in [1:N]$.

Exercise 8.12. Fixing a matrix $H \in \mathbb{R}^{R \times N}$, let $\varphi^H : \mathbb{R}^D \to \mathbb{R}$ with $D := KR$ be defined by $\varphi^H(W) := \|V - WH\|^2$ for $W \in \mathbb{R}^{K \times R}$ (see (8.80)). Compute the gradient

of φ^H (similar to the calculation of the gradient of φ^H in (8.73) to (8.78)). From this, derive the update rule as specified in (8.81).

Solution to Exercise 8.12. Let $W_{\kappa\rho}$ for $\kappa \in [1:K]$ and $\rho \in [1:R]$ denote the variables of the function φ^H. The partial derivatives of φ^H with regard to the variables $W_{\kappa\rho}$ are computed as follows:

$$\begin{aligned}
\frac{\partial \varphi^H}{\partial W_{\kappa\rho}} &= \frac{\partial \left(\sum_{k=1}^{K} \sum_{n=1}^{N} \left(V_{kn} - \sum_{r=1}^{R} W_{kr} H_{rn} \right)^2 \right)}{\partial W_{\kappa\rho}} \\
&= \frac{\partial \left(\sum_{n=1}^{N} \left(V_{\kappa n} - \sum_{r=1}^{R} W_{\kappa r} H_{rn} \right)^2 \right)}{\partial W_{\kappa\rho}} \\
&= \sum_{n=1}^{N} 2 \Big(V_{\kappa n} - \sum_{r=1}^{R} W_{\kappa r} H_{rn} \Big) \cdot (-H_{\rho n}) \\
&= 2 \Big(\sum_{r=1}^{R} \sum_{n=1}^{N} W_{\kappa r} H_{rn} H_{\rho n} - \sum_{n=1}^{N} V_{\kappa n} H_{\rho n} \Big) \\
&= 2 \Big(\sum_{r=1}^{R} W_{\kappa r} \sum_{n=1}^{N} \Big(H_{rn} H_{n\rho}^{\top} \Big) - \sum_{n=1}^{N} V_{\kappa n} H_{n\rho}^{\top} \Big) \\
&= 2 \big((WHH^{\top})_{\kappa\rho} - (VH^{\top})_{\kappa\rho} \big)
\end{aligned}$$

Starting with an initial guess $W^{(0)} \in \mathbb{R}^{K \times R}$, one obtains the following additive update rules:

$$W_{kr}^{(\ell+1)} = W_{kr}^{(\ell)} - \gamma_{kr}^{(\ell)} \cdot \Big(\big(W^{(\ell)} H H^{\top} \big)_{kr} - \big(V H^{\top} \big)_{kr} \Big)$$

for $\ell = 0, 1, 2, \ldots$ and some suitable parameters $\gamma_{kr}^{(\ell)} \geq 0$.

Exercise 8.13. Show that, in the case of a "perfect" factorization $V = WH$, the matrices W and H are a fixed point of the multiplicative update rules (8.83) and (8.85).

Solution to Exercise 8.13. Let $V = WH$. Then, using W and H in (8.83), one obtains

$$H_{rn} \cdot \frac{\big(W^{\top} V \big)_{rn}}{\big(W^{\top} W H \big)_{rn}} = H_{rn} \cdot \frac{\big(W^{\top} V \big)_{rn}}{\big(W^{\top} V \big)_{rn}} = H_{rn}$$

for $r \in [1:R]$ and $n \in [1:N]$. Similarly, using W and H in (8.85), one obtains

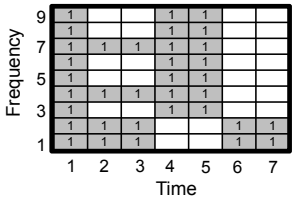
$$W_{kr} \cdot \frac{\big(V H^{\top} \big)_{kr}}{\big(W H H^{\top} \big)_{kr}} = W_{kr} \cdot \frac{\big(V H^{\top} \big)_{kr}}{\big(V H^{\top} \big)_{kr}} = W_{kr}$$

for $k \in [1:K]$ and $r \in [1:R]$. This shows that H and W are fixed points of the multiplicative update rules.

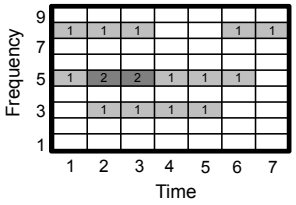
Exercise 8.14. Let V be a $(K \times N)$ matrix with nonnegative entries. As in Figure 8.20a, we consider in this exercise an exact matrix factorization $V = WH$

with a nonnegative $(K \times R)$ matrix W and a nonnegative $(R \times N)$ matrix H. In the following examples, we have $N = 7$ and $K = 9$. Determine for each of the two matrices at least two decompositions $V = WH$ using $R = 3$:

(a)



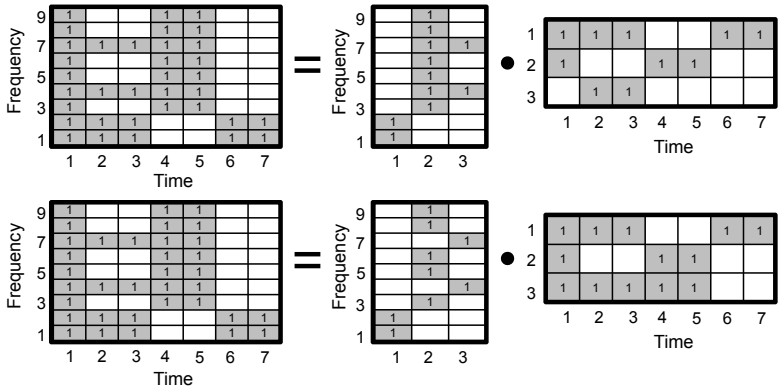
(b)



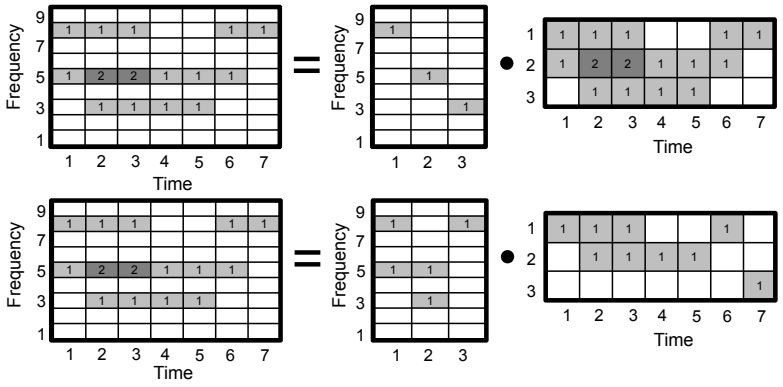
Explain why in these examples there are no exact factorizations when using $R = 2$.

Solution to Exercise 8.14.

(a)



(b)



In both examples, the matrix V has a rank of three. Therefore, it is not possible to factorize these matrices using $R = 2$.

Meinard Müller

Fundamentals of Music Processing

Audio, Analysis, Algorithms, Applications

Exercises and Solutions