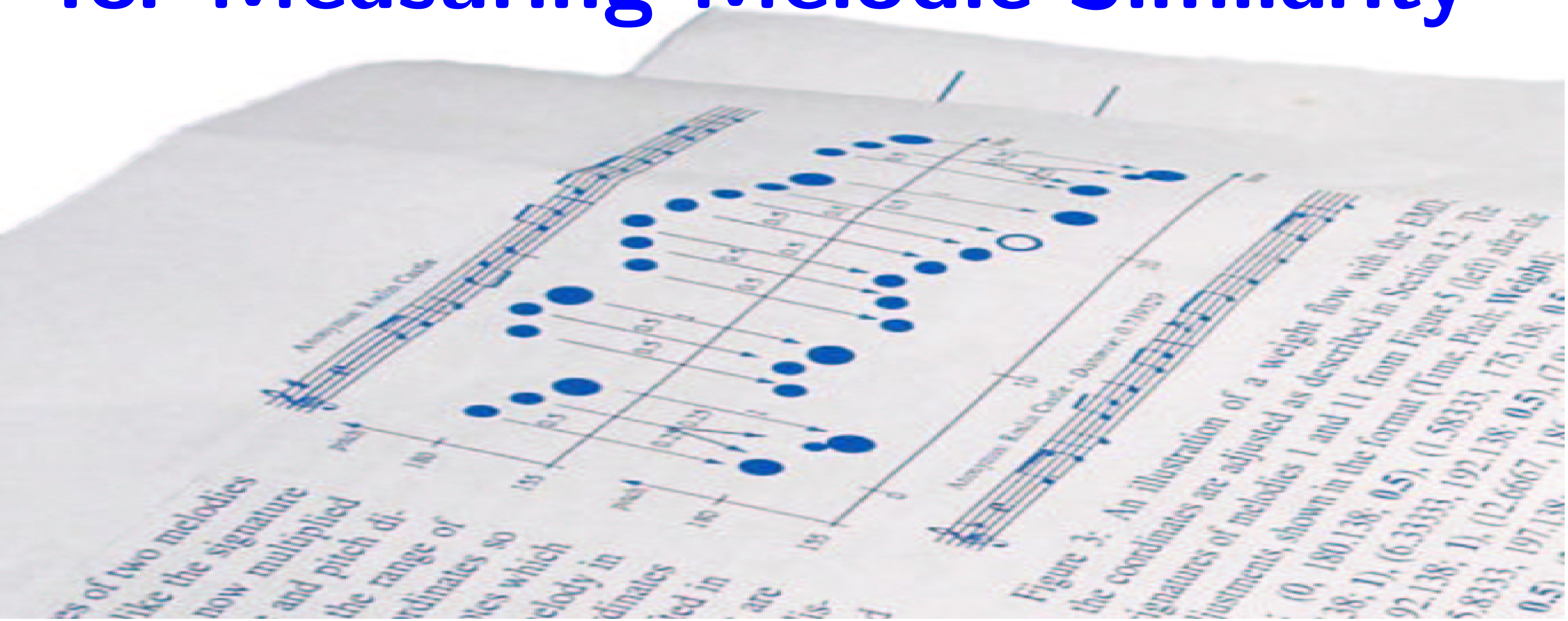
Using Transportation Distances for Measuring Melodic Similarity



Rainer Typke, Panos Giannopoulos, Remco C. Veltkamp, Frans Wiering, René van Oostrum

Utrecht University, Institute of Information and Computing Sciences
Center for Geometry, Imaging and Virtual Environments

Why transportation distances are promising

Using transportation distances such as the Earth Mover's Distance (EMD) for comparing symbolic music notation seems to be a good idea:

- Good matching results
- Efficient search possible (e. g. with Proportional Transportation Distance)
- Polyphonic searches in polyphonic music pose no additional complications in comparison to monophonic matching
- Can be easily adjusted to different purposes by modifying weighting scheme and ground distance
- Opens up interesting possibilities (e. g., QBH without separate note onset detection step)

The EMD has been used for comparing audio, but we are not aware of previous work involving the comparison of notated music.

Good matching results

In a database containing 476,000 melodies, the top 17 matches contain 12 out of 15 known occurrences of the query, “Roslin Castle”.

Distance measure: EMD
Weights: Duration only

Comparison with earlier work involving the RISM A/II data

Grouping occurrences of “Roslin Castle” together

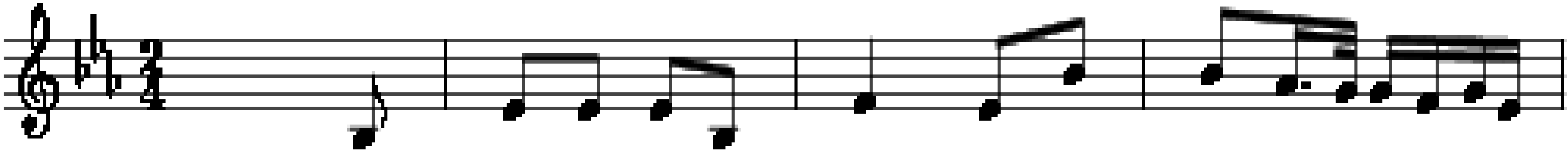
- Howard (1998) encoded the RISM A/II collection in the DARMS format and tried various sorting methods. None of his methods sorted more than 46 % of the known occurrences together.
- We were able to group 73 % together.

Identifying Anonymous Pieces

- Schlichte (1990) was able to identify 2.08 % of anonymous pieces in the RISM A/II collection by looking for identical Plaine & Easie encodings.
- We compared about 80,000 anonymous incipits to all 476,000 pieces in our database and could identify 3.9 %.

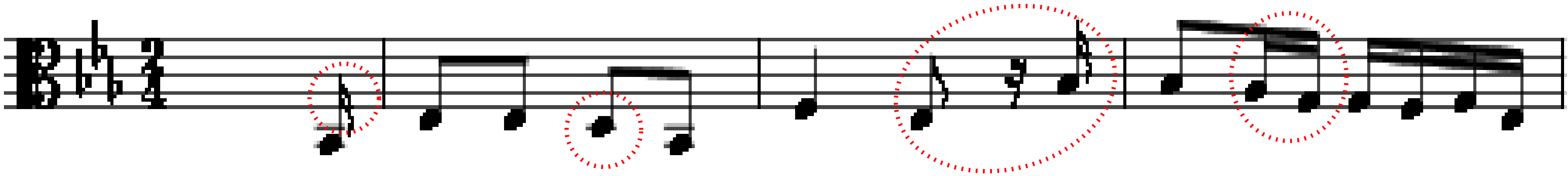
Example of an identified piece

Query: Anonymus: Andante. Keyboard piece without title in a collection of manuscripts with piano pieces by Clementi, J. Chr. Bach, Boccherini, and Pleyel, Musikbibliothek, Kloster Einsiedeln, Switzerland.

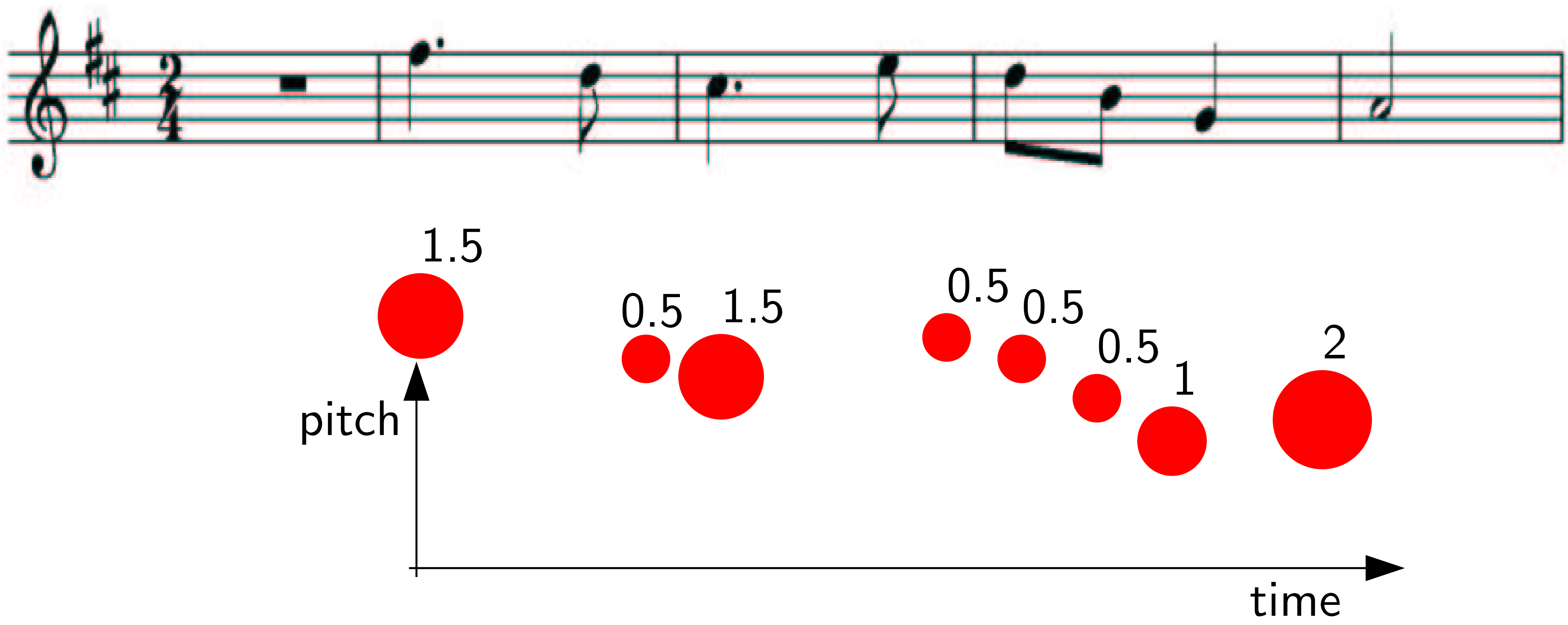


Match: Ignaz Umlauff/ A. F. J. Eberl/W. A. Mozart: Ariette variée for piano. Excerpt from "Irrlicht", "Zu Steffen sprach im Traume." Manuscripts in Brescia and Dubrovnik.

Match: I. Umlauff: Singspiel "Das Irrlicht", Basso: "Zu Steffen sprach im Traume" (Due corni, due fagotti, due violini, due viole e basso), manuscript in Valdemars Slot, Tåsinge, Denmark



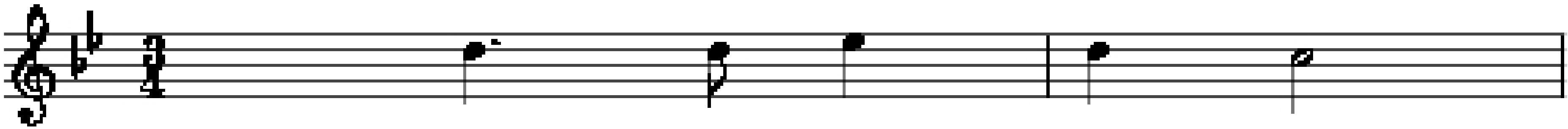
17,895 more examples on http://give-lab.cs.uu.nl/MIR/anon/idx.html



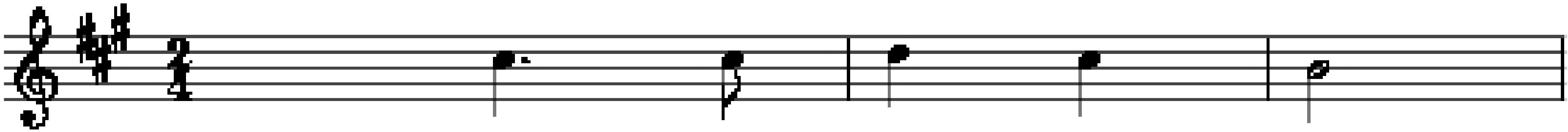
- Coordinates represent note onset time and pitch.
- Weights should reflect the notes' importance. So far, we used mainly the duration, but other aspects can also be reflected in the weights.

Example for Weight Components: Stress Weight

Jean-Baptiste Lully: La Grotte de Versailles



Anonymus: Litanies (Coro, without title)



- These melodies differ only in the measure structure.
- By adding a weight component for emphasized notes in every bar, the measure structure can be taken into account and a distance > 0 can be achieved for cases like this.

The Earth Mover's Distance

Measures the minimum amount of work needed to transform one weighted point set into the other by moving weight.

The set of all possible flows is defined by these constraints:

- no negative flow component.
- no point emits or receives more weight than it has.
- the lighter of the two point sets is completely matched.

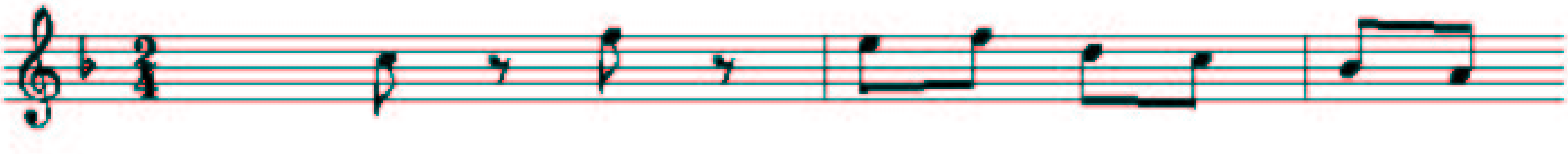
The EMD is the weighted sum of the optimum flow components' distances, divided by the matched weight:

$$\mathrm{EMD}(A, B) = \frac{\min_{F \in \mathcal{F}} \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} d_{ij}}{\min(W, U)}$$

EMD: An example

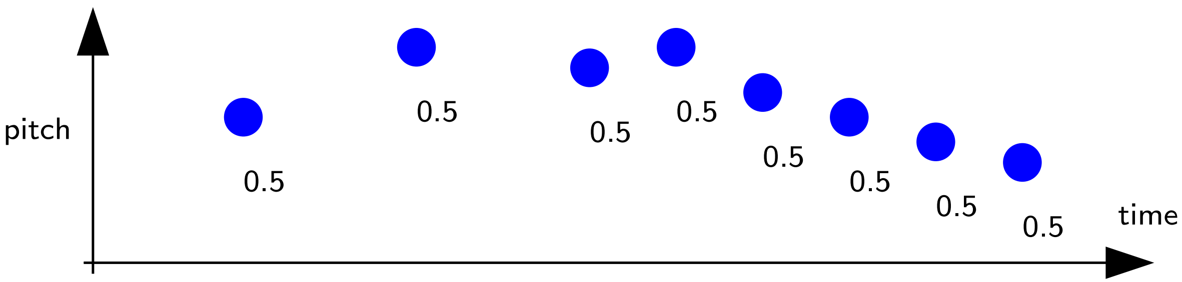
Anonymus: Roslin Castle

Joseph Aloys Schmittbauer: Lauda Sion – Distance: 0.79



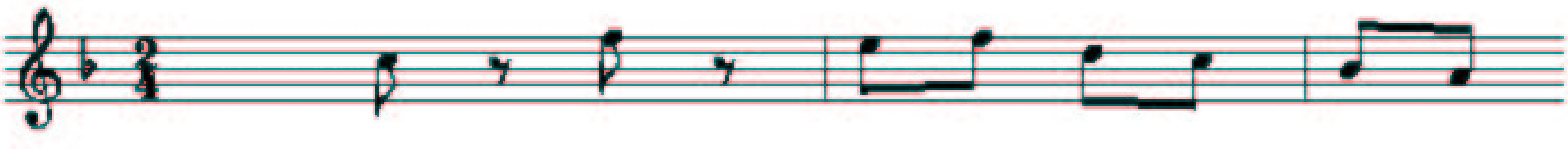
EMD: An example

Anonymus: Roslin Castle



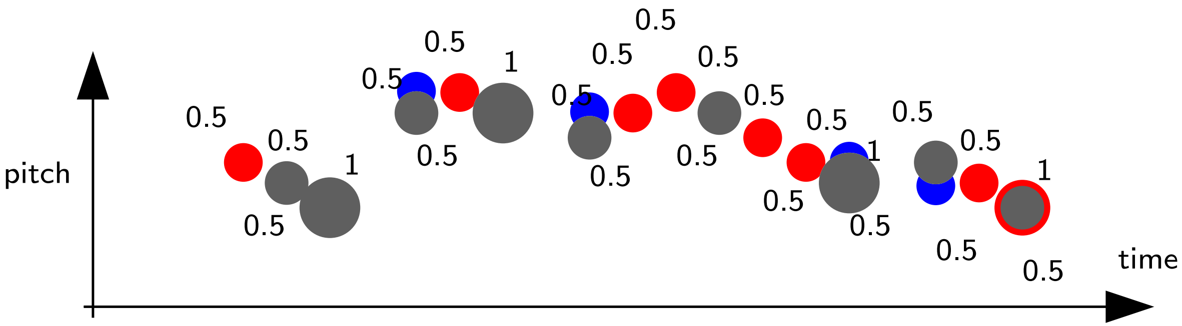
blue: Lauda Sion.

Joseph Aloys Schmittbauer: Lauda Sion – Distance: 0.79



EMD: An example

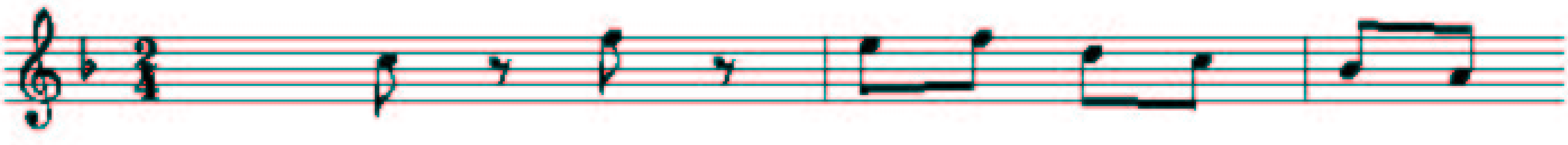
Anonymus: Roslin Castle



red/gray: Roslin Castle. Unmatched points or parts thereof are shown in gray.

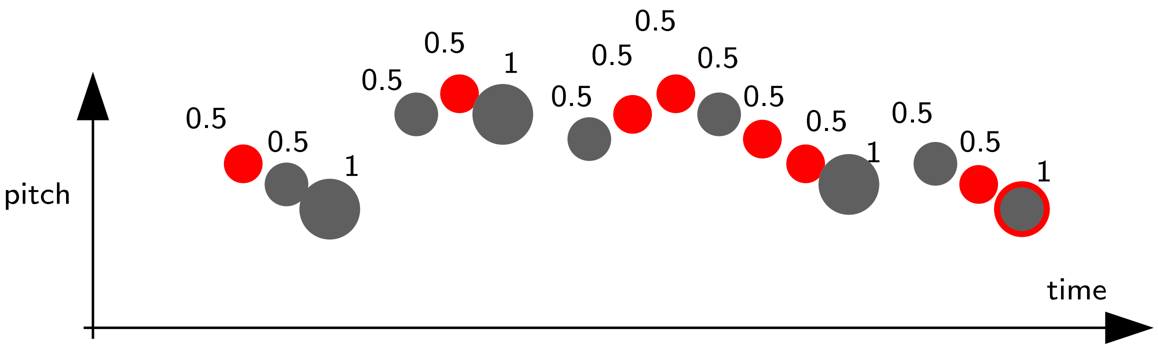
blue: Lauda Sion.

Joseph Aloys Schmittbauer: Lauda Sion – Distance: 0.79



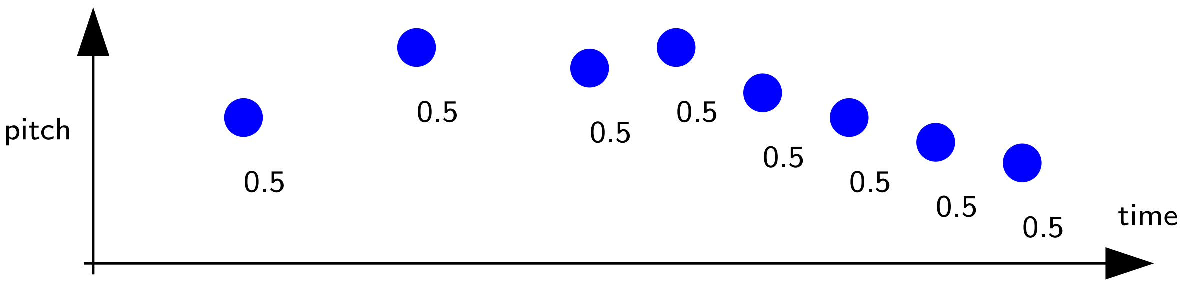
EMD: An example

Anonymus: Roslin Castle

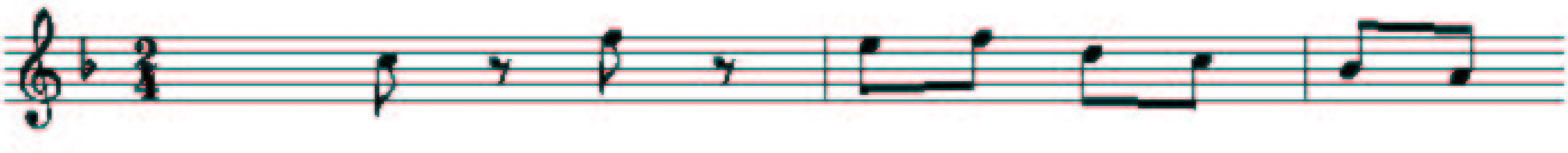


red/gray: Roslin Castle. Unmatched points or parts thereof are shown in gray.

blue: Lauda Sion.

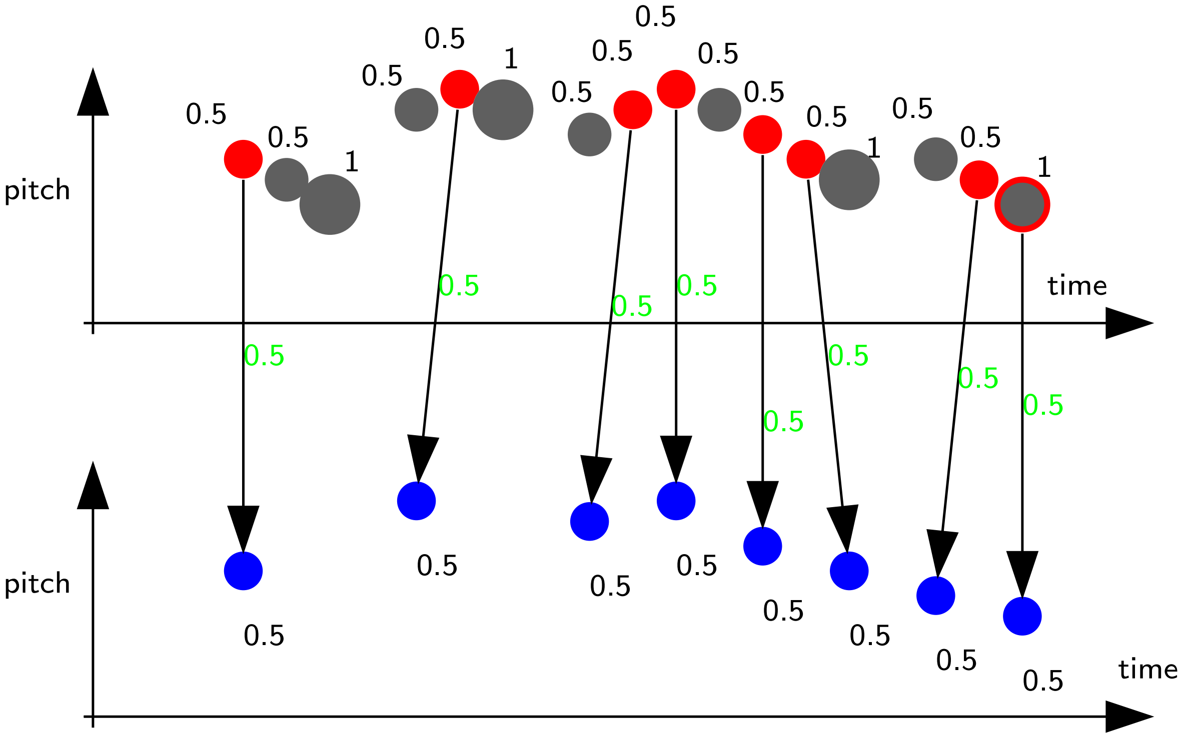


Joseph Aloys Schmittbauer: Lauda Sion – Distance: 0.79



EMD: An example

Anonymus: Roslin Castle

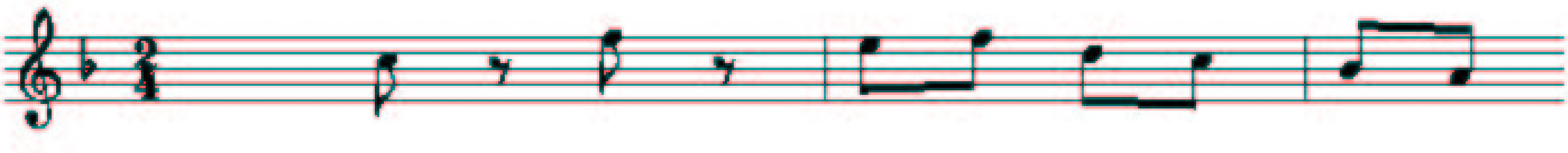


red/gray: Roslin Castle. Unmatched points or parts thereof are shown in gray.

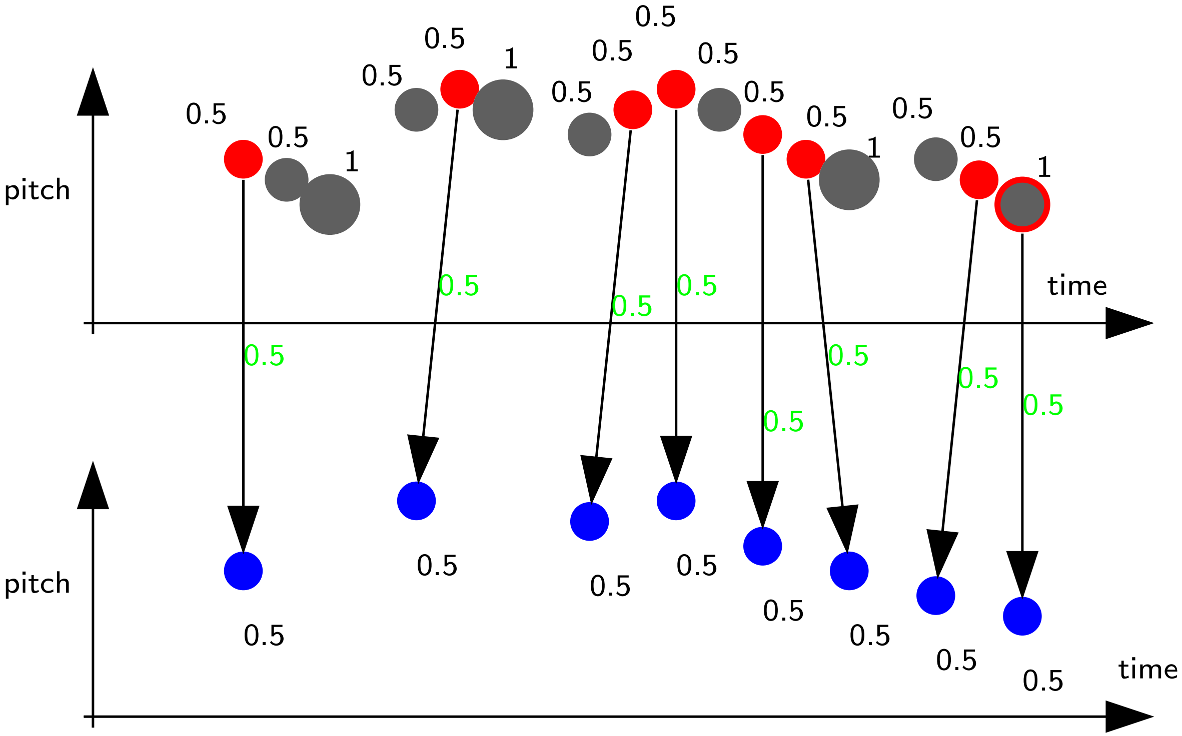
blue: Lauda Sion.

Weights are shown as black numbers, flows as green numbers.

Joseph Aloys Schmittbauer: Lauda Sion – Distance: 0.79



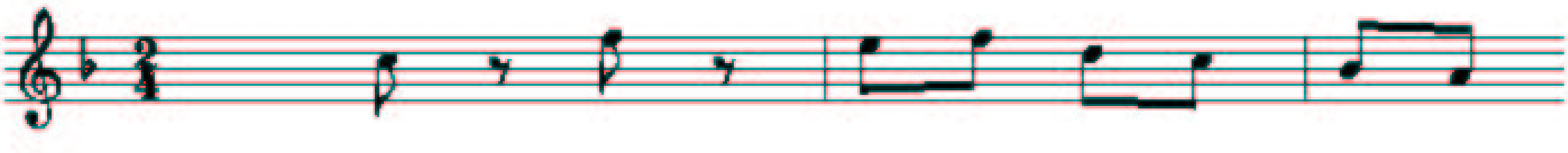
Anonymus: Roslin Castle



i	j	f_{ij}	d_{ij}
1	1	0.5	0
5	2	0.5	1.5
8	3	0.5	1.5
9	4	0.5	0
11	5	0.5	0
12	6	0.5	1.5
15	7	0.5	1.789
16	8	0.5	0

$$\text{EMD} = \frac{3(0.5 \cdot 1.5) + 0.5 \cdot 1.789}{8 \cdot 0.5} = 0.79$$

Joseph Aloys Schmittbauer: Lauda Sion – Distance: 0.79



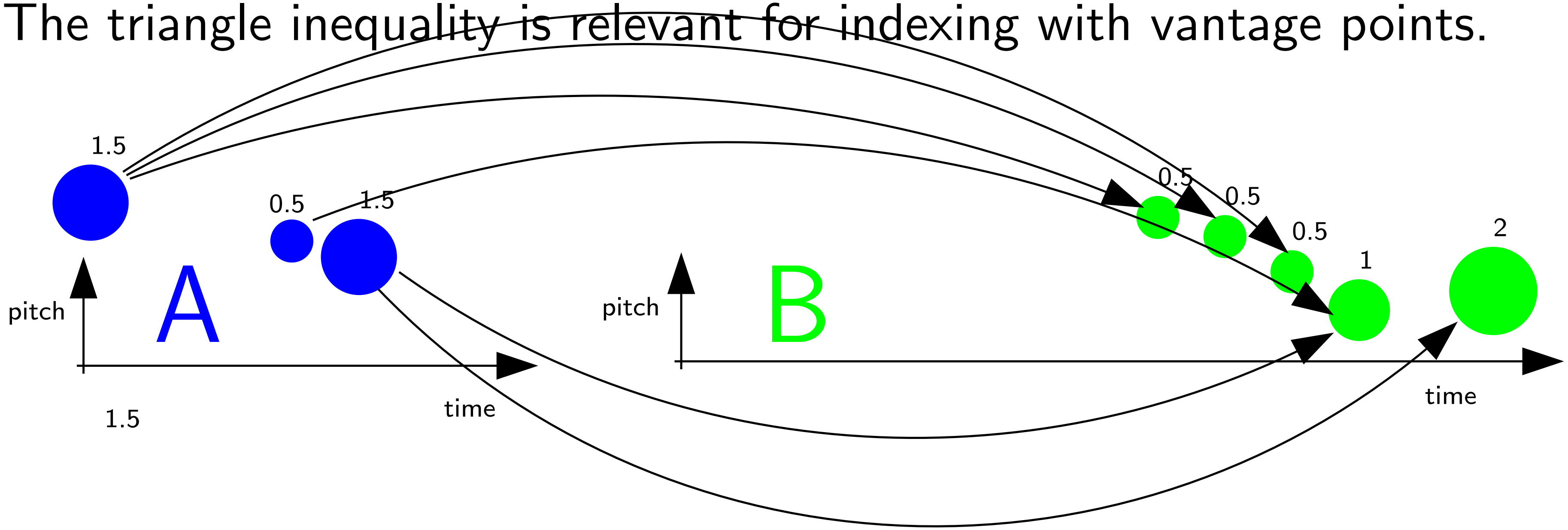
Properties of the EMD

- The EMD is continuous.
- For unequal weight sums, it does not have the positivity property. I. e., partial matching is possible, and there are cases where the EMD does not distinguish between different pairs of non-identical sets.
- For unequal weight sums, the EMD does not obey the triangle inequality.

The triangle inequality is relevant for indexing with vantage points.

- The EMD is continuous.
- For unequal weight sums, it does not have the positivity property. I. e., partial matching is possible, and there are cases where the EMD does not distinguish between different pairs of non-identical sets.
- For unequal weight sums, the EMD does not obey the triangle inequality.

The triangle inequality is relevant for indexing with vantage points.

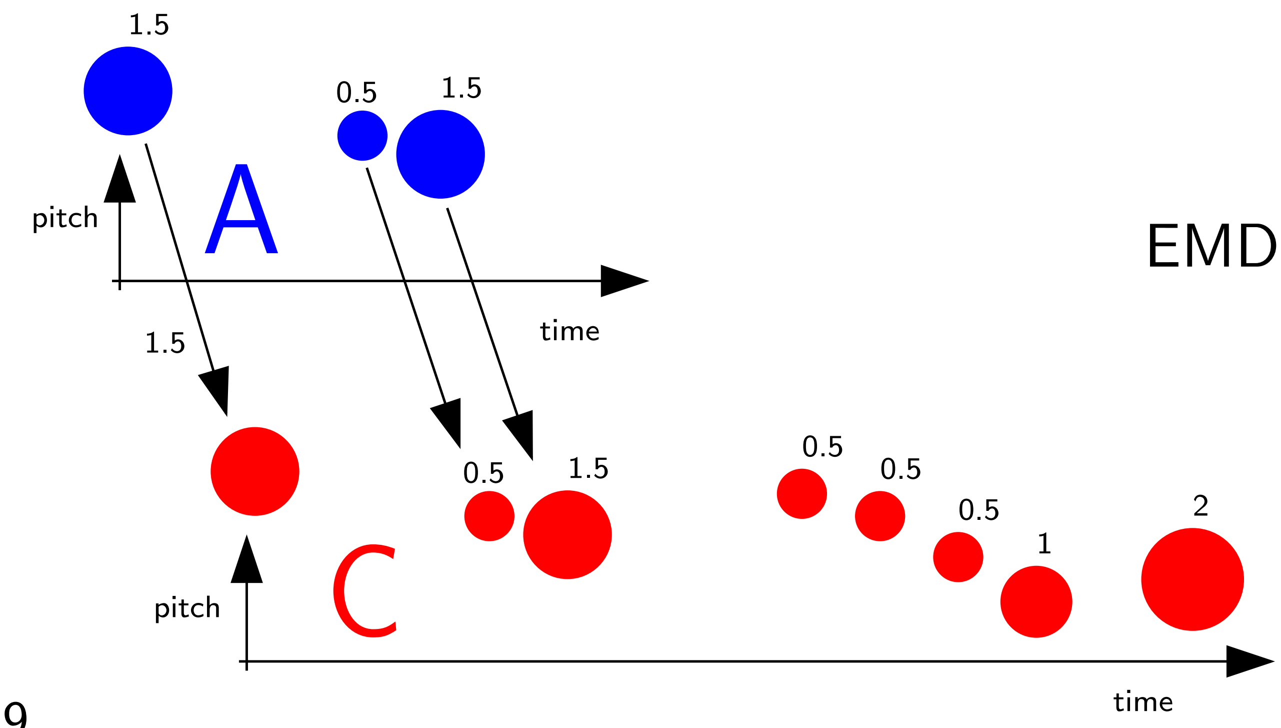


EMD(A,B) > 0

Properties of the EMD

- The EMD is continuous.
- For unequal weight sums, it does not have the positivity property. I. e., partial matching is possible, and there are cases where the EMD does not distinguish between different pairs of non-identical sets.
- For unequal weight sums, the EMD does not obey the triangle inequality.

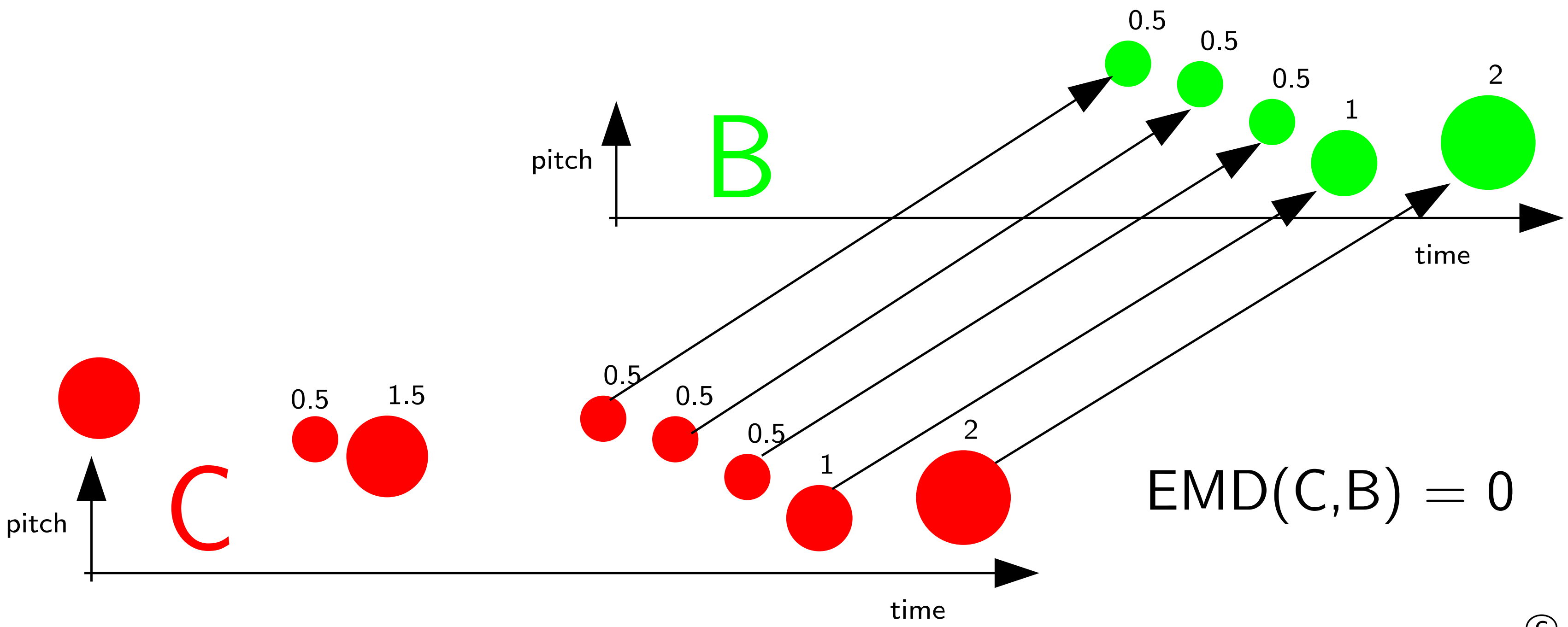
The triangle inequality is relevant for indexing with vantage points.



EMD(A,C) = 0

- The EMD is continuous.
- For unequal weight sums, it does not have the positivity property. I. e., partial matching is possible, and there are cases where the EMD does not distinguish between different pairs of non-identical sets.
- For unequal weight sums, the EMD does not obey the triangle inequality.

The triangle inequality is relevant for indexing with vantage points.

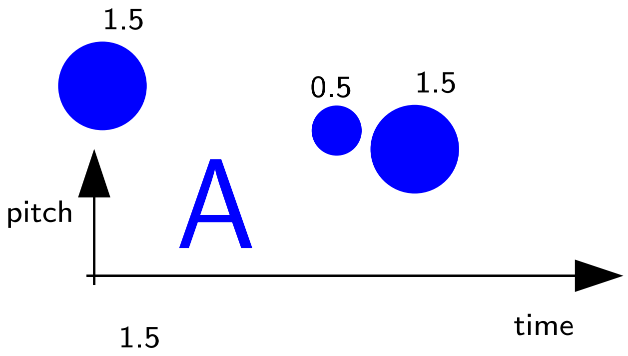


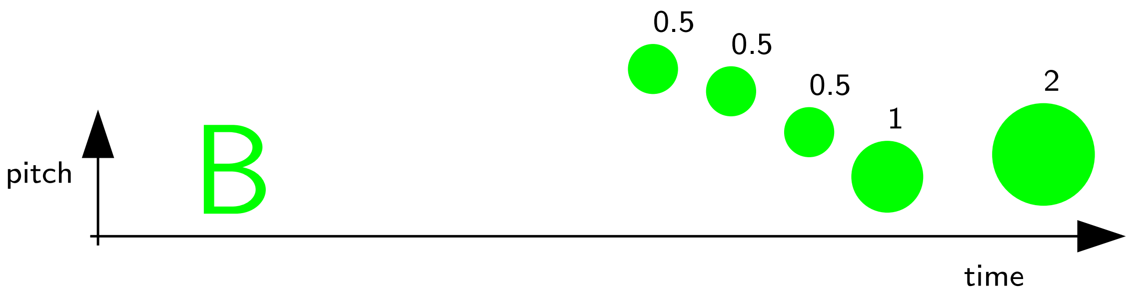
$\mathrm{EMD(C,B)} = 0$

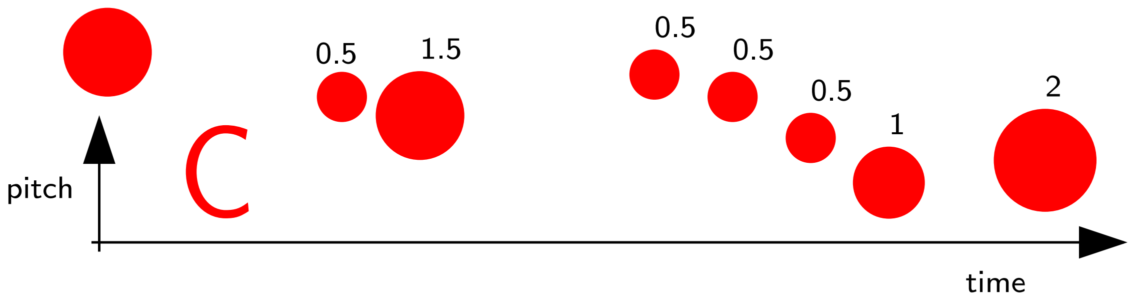
Properties of the EMD

- The EMD is continuous.
- For unequal weight sums, it does not have the positivity property. I. e., partial matching is possible, and there are cases where the EMD does not distinguish between different pairs of non-identical sets.
- For unequal weight sums, the EMD does not obey the triangle inequality.

The triangle inequality is relevant for indexing with vantage points.





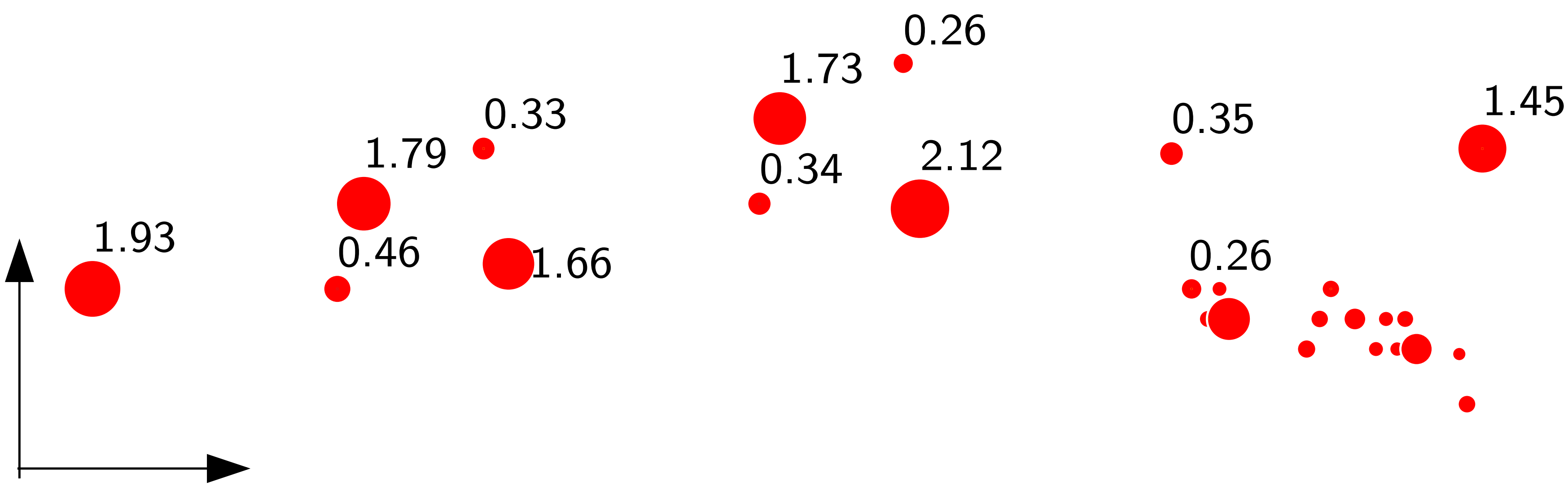


$$\text{EMD(A,B)} > \text{EMD(A,C)} + \text{EMD(C, B)}$$

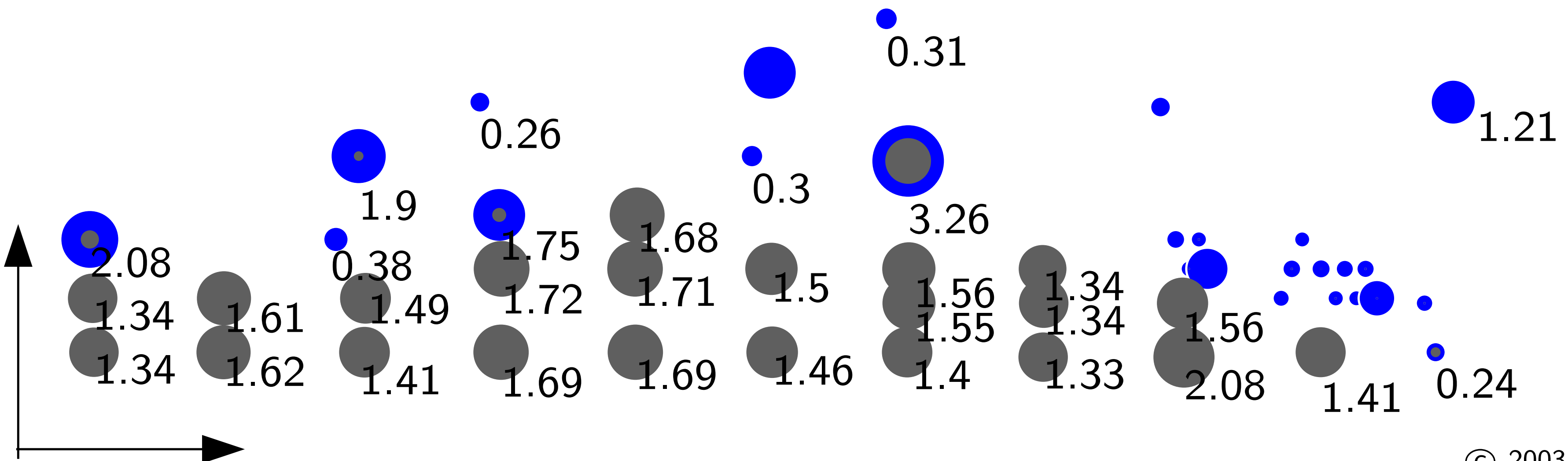
Partial matching with the EMD in a polyphonic piece

Partial matching with the EMD in a polyphonic piece

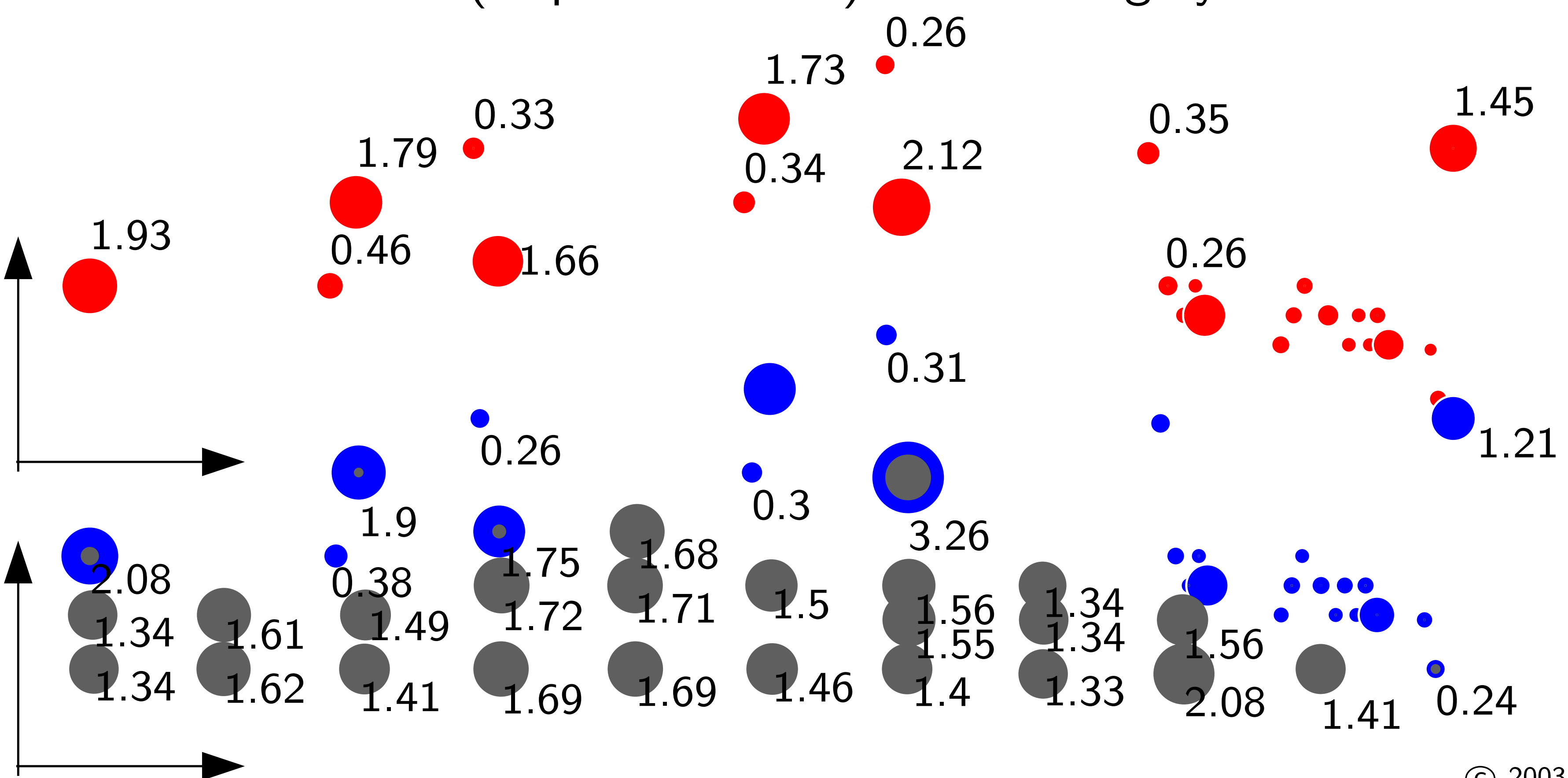
red: top voice only (monophonic), recorded with a MIDI keyboard



blue/gray: all voices, in a different MIDI keyboard recording. The non-matched notes (or parts thereof) are shown gray.

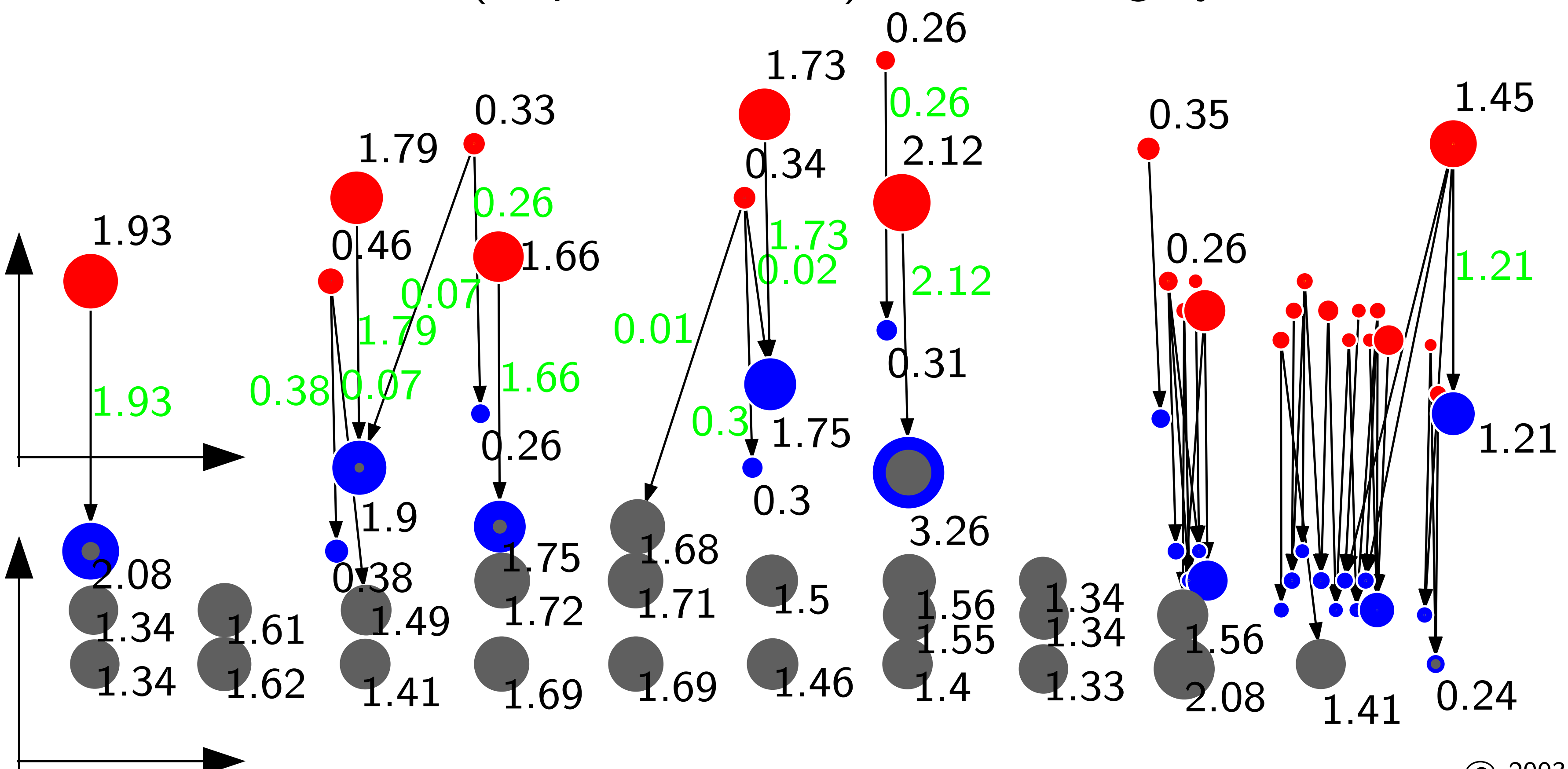


red: top voice only (monophonic), recorded with a MIDI keyboard
blue/gray: all voices, in a different MIDI keyboard recording. The non-matched notes (or parts thereof) are shown gray.



Partial matching with the EMD in a polyphonic piece

red: top voice only (monophonic), recorded with a MIDI keyboard
blue/gray: all voices, in a different MIDI keyboard recording. The non-matched notes (or parts thereof) are shown gray.



The Proportional Transportation Distance (Giannopoulos & Veltkamp 2002)

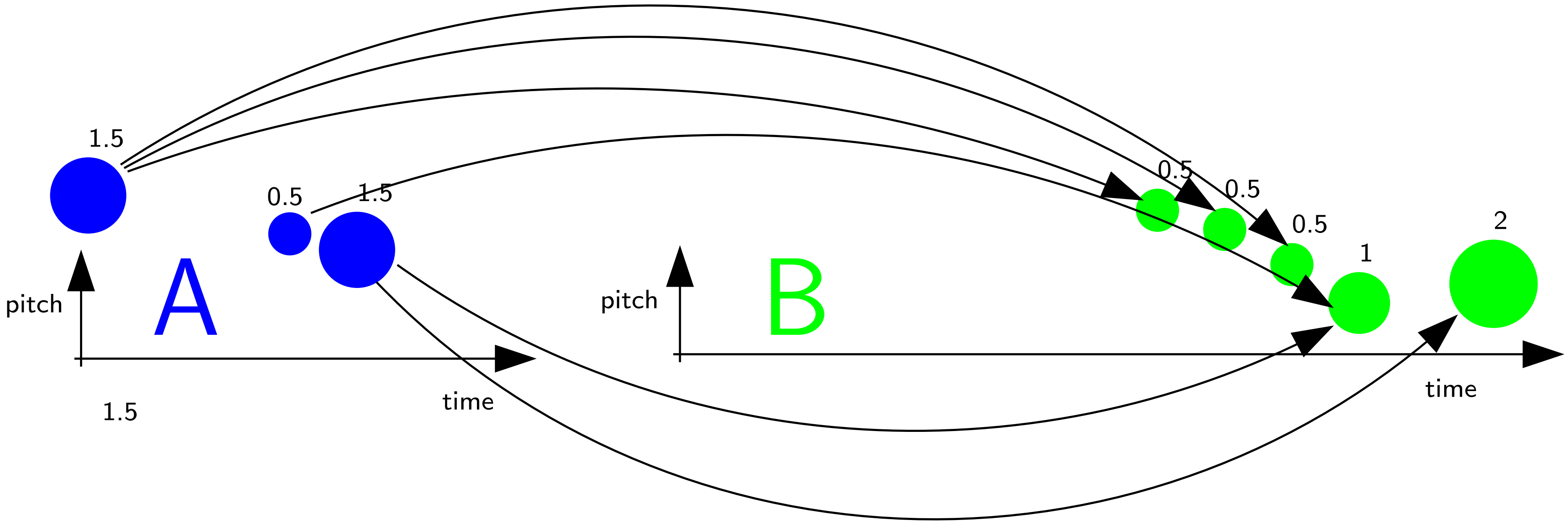
- Takes a weight surplus into account.
- Triangle inequality holds.
- Still does not have the positivity property.

Calculation: for both point sets, the total weight is normalized to 1, while preserving the weight proportions. Then, the EMD is calculated.

The Proportional Transportation Distance (Giannopoulos & Veltkamp 2002)

- Takes a weight surplus into account.
- Triangle inequality holds.
- Still does not have the positivity property.

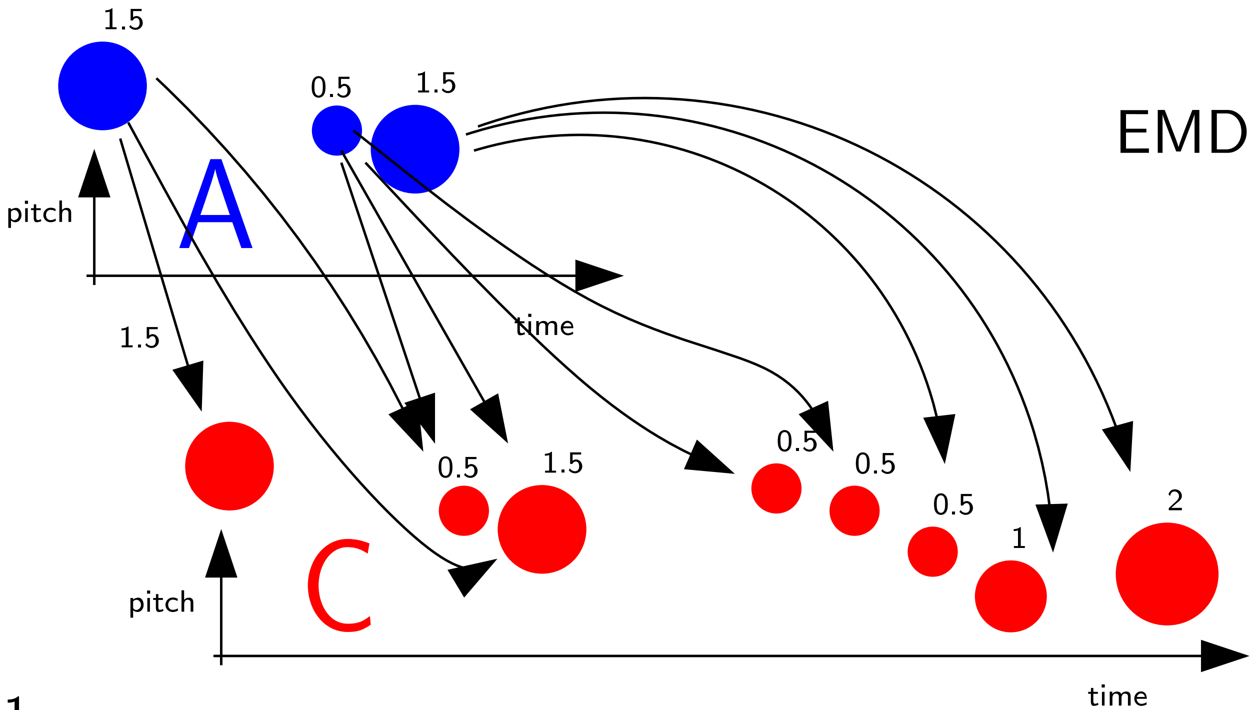
Calculation: for both point sets, the total weight is normalized to 1, while preserving the weight proportions. Then, the EMD is calculated.



EMD(A,B) $>$ 0

- Takes a weight surplus into account.
- Triangle inequality holds.
- Still does not have the positivity property.

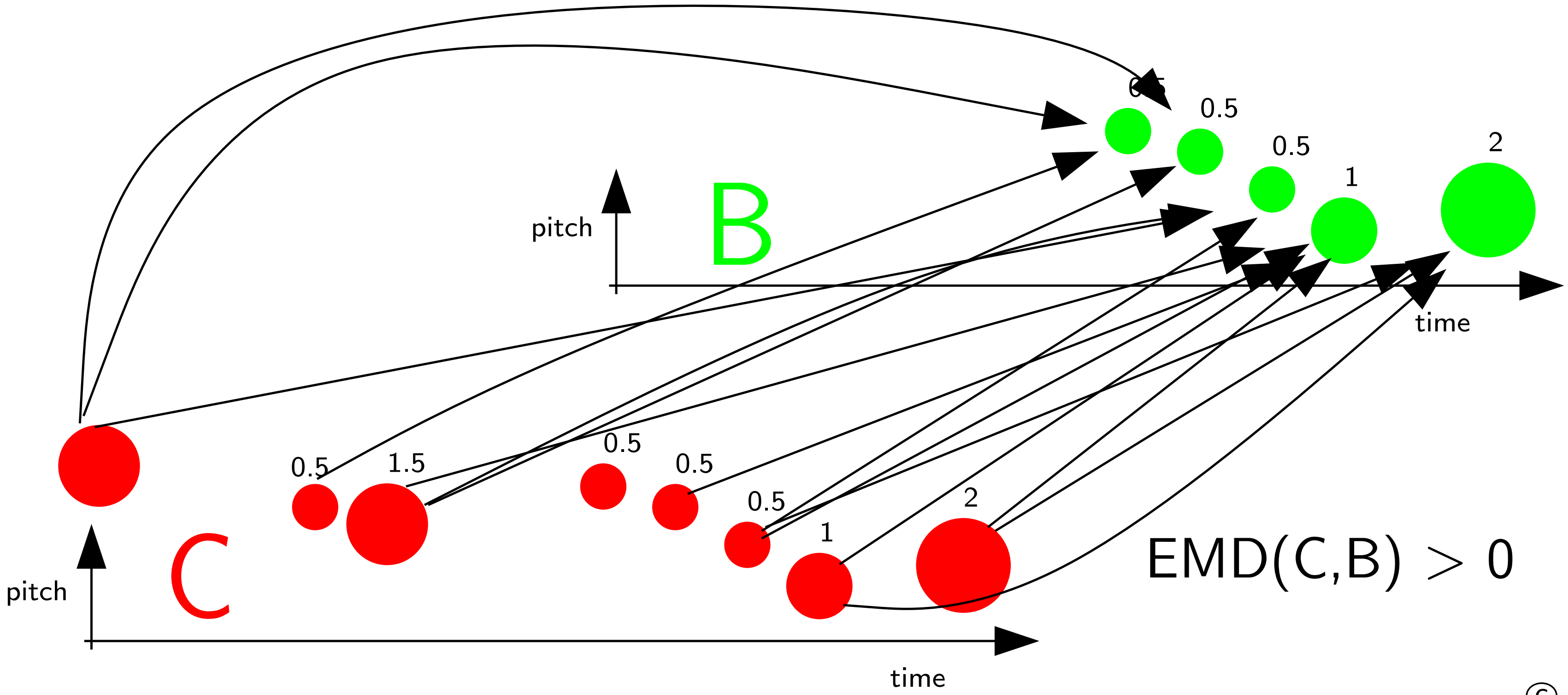
Calculation: for both point sets, the total weight is normalized to 1, while preserving the weight proportions. Then, the EMD is calculated.



EMD(A,C) $>$ 0

- Takes a weight surplus into account.
- Triangle inequality holds.
- Still does not have the positivity property.

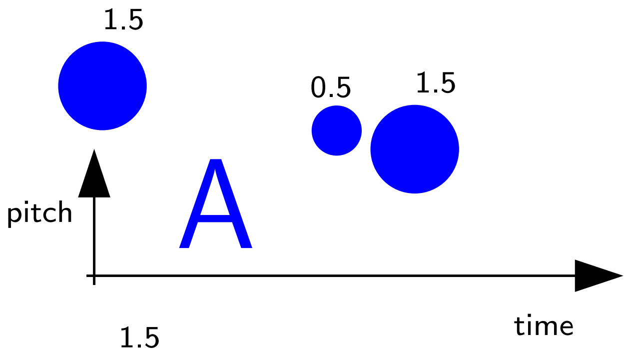
Calculation: for both point sets, the total weight is normalized to 1, while preserving the weight proportions. Then, the EMD is calculated.

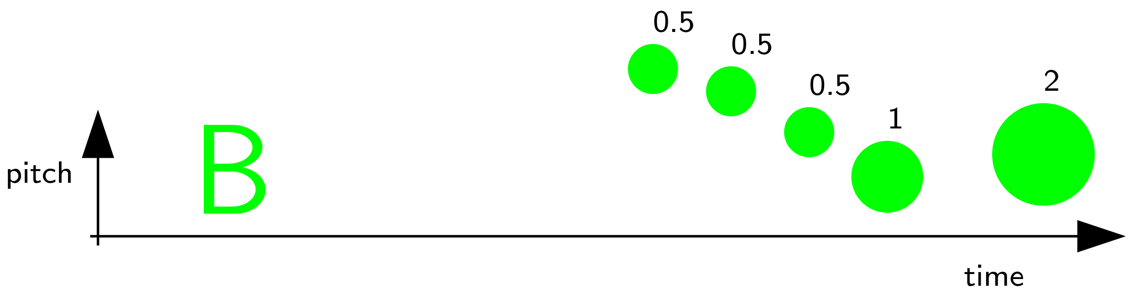


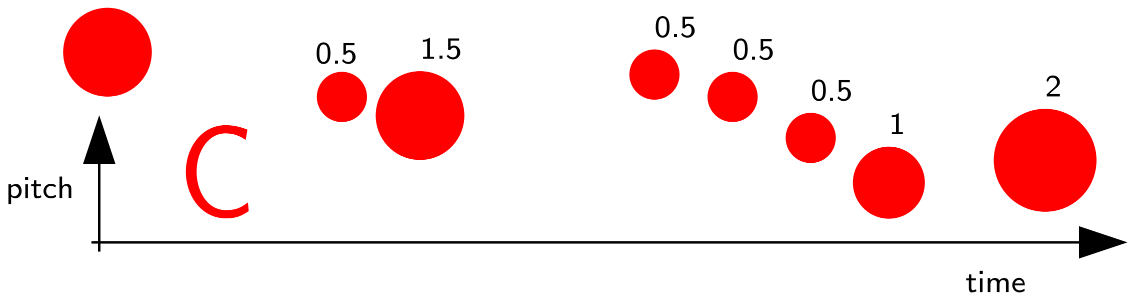
The Proportional Transportation Distance (Giannopoulos & Veltkamp 2002)

- Takes a weight surplus into account.
- Triangle inequality holds.
- Still does not have the positivity property.

Calculation: for both point sets, the total weight is normalized to 1, while preserving the weight proportions. Then, the EMD is calculated.



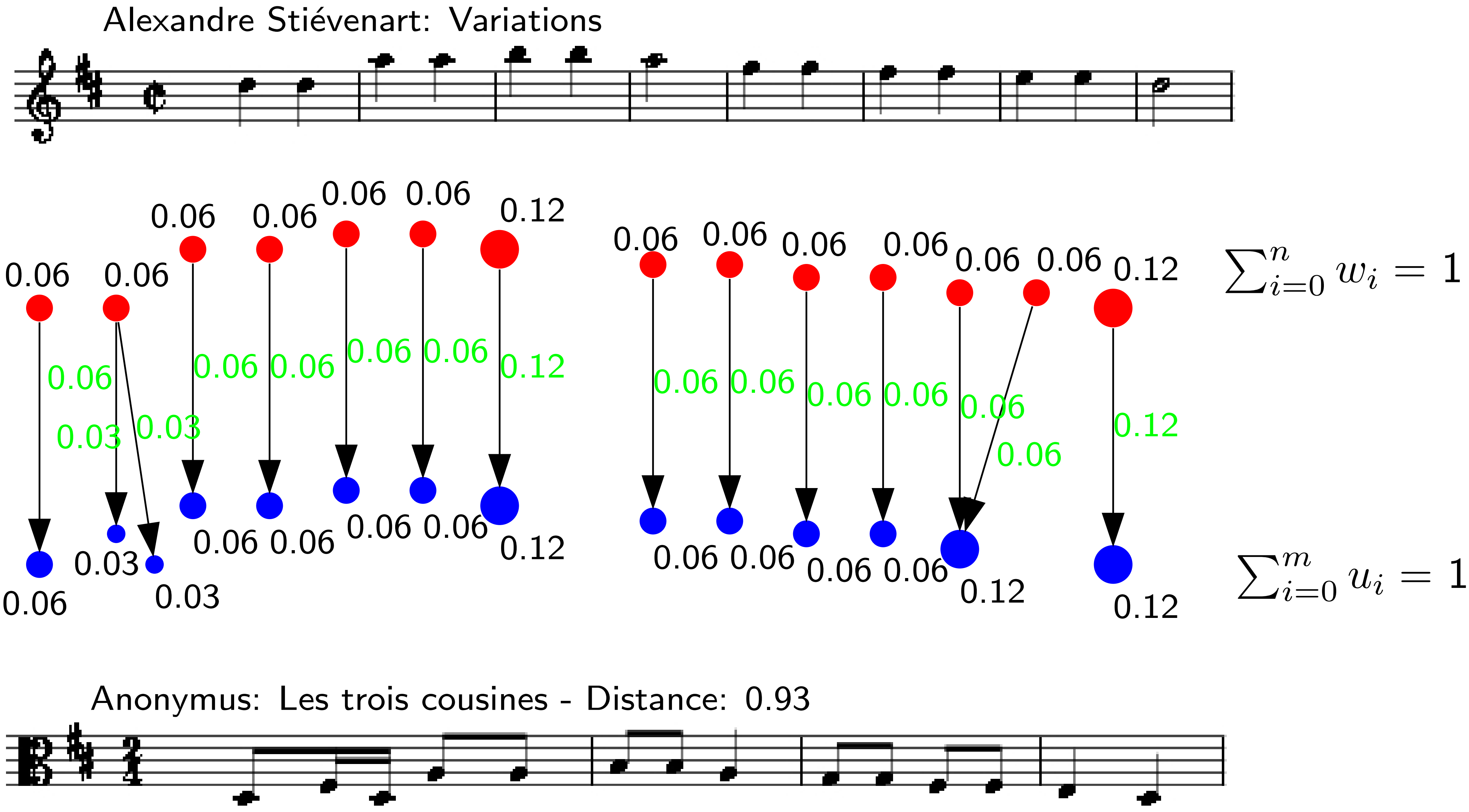




$$\text{EMD}(A,B) <= \text{EMD}(A,C) + \text{EMD}(C,B)$$

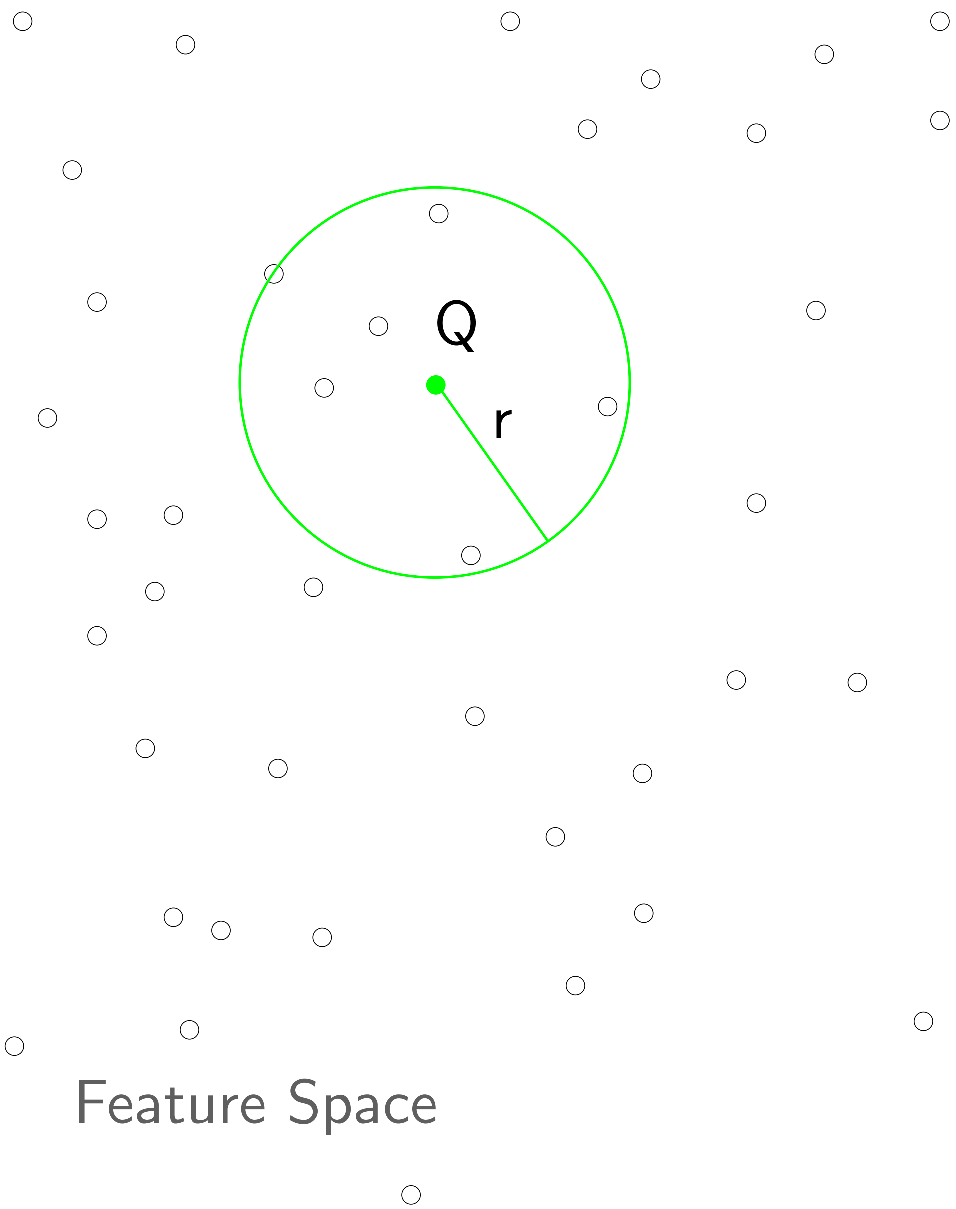
PTD Matching example

- The weight sum is 1 for both melodies.
- Because of this, augmentation or diminution is ignored.



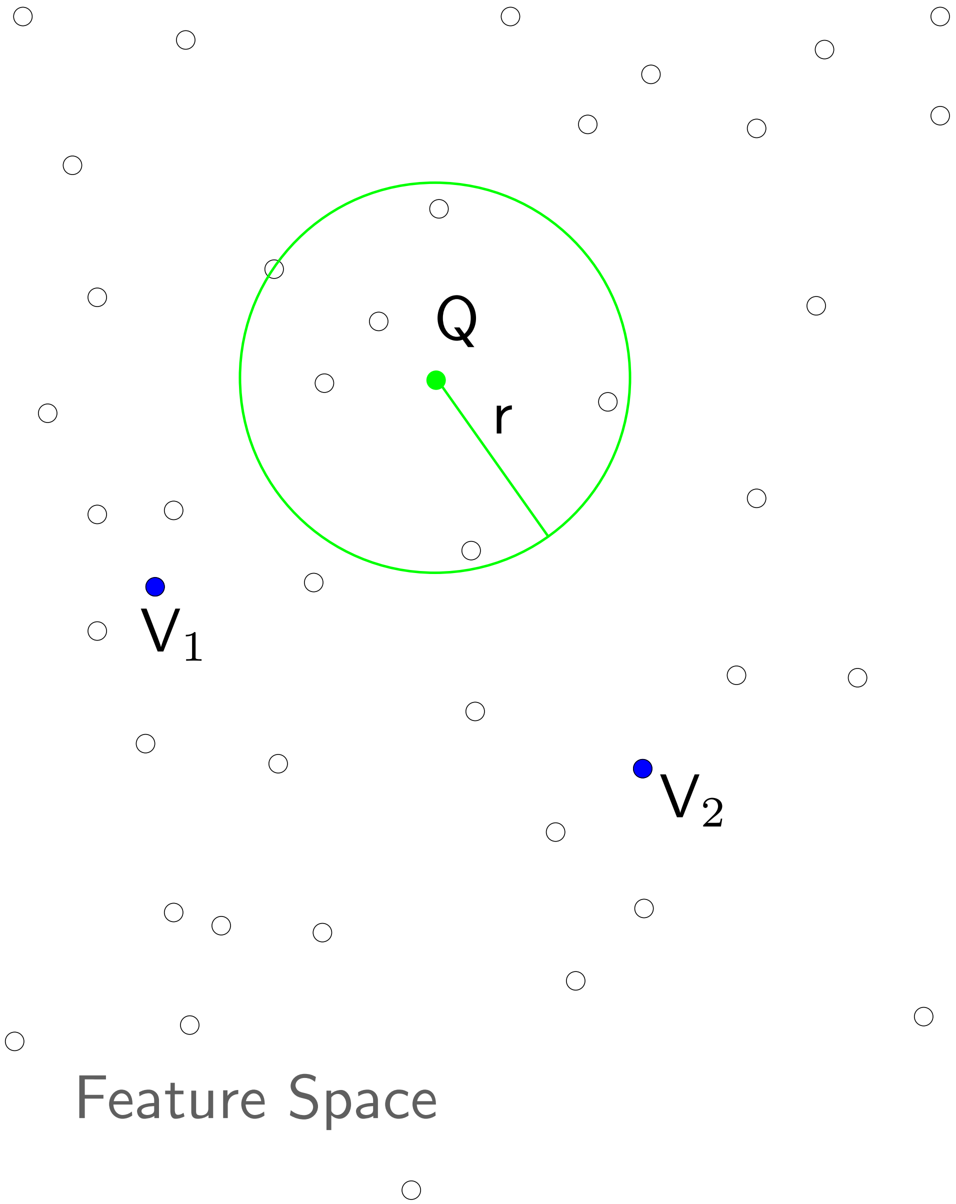
Indexing with Vantage Objects

We want to retrieve all objects with a distance <= r from the query object Q.



Indexing with Vantage Objects

We want to retrieve all objects with a distance $<=$ r from the query object Q.

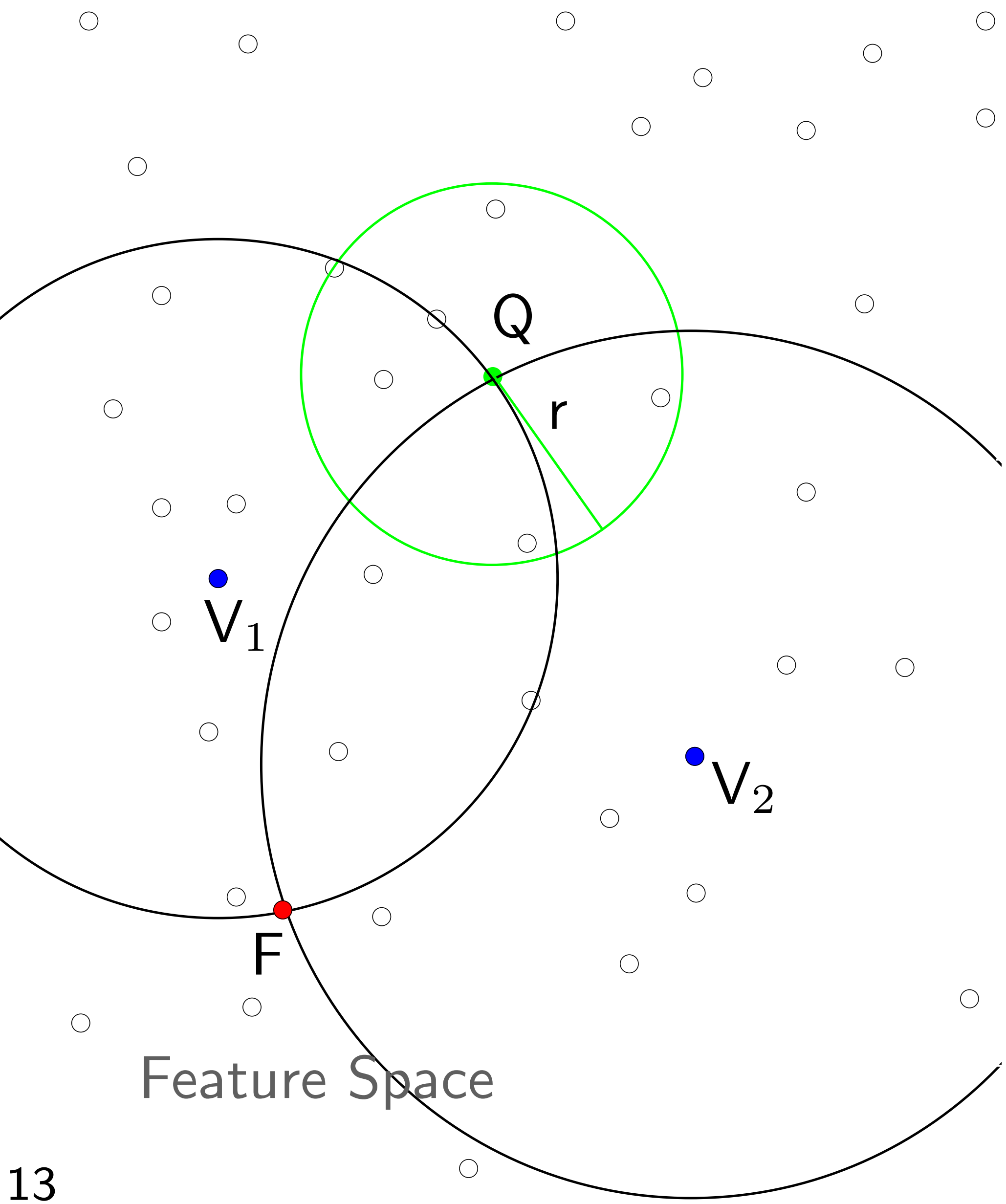


V_1, V_2: Vantage objects.

Instead of searching the whole feature space:

- For each object, pre-calculate the distances to vantage objects.
- Use these distances as coordinates in a Euclidean space.
- Restrict the search to those objects with a Euclidean distance $<=$ r in the Euclidean space of vantage distances.

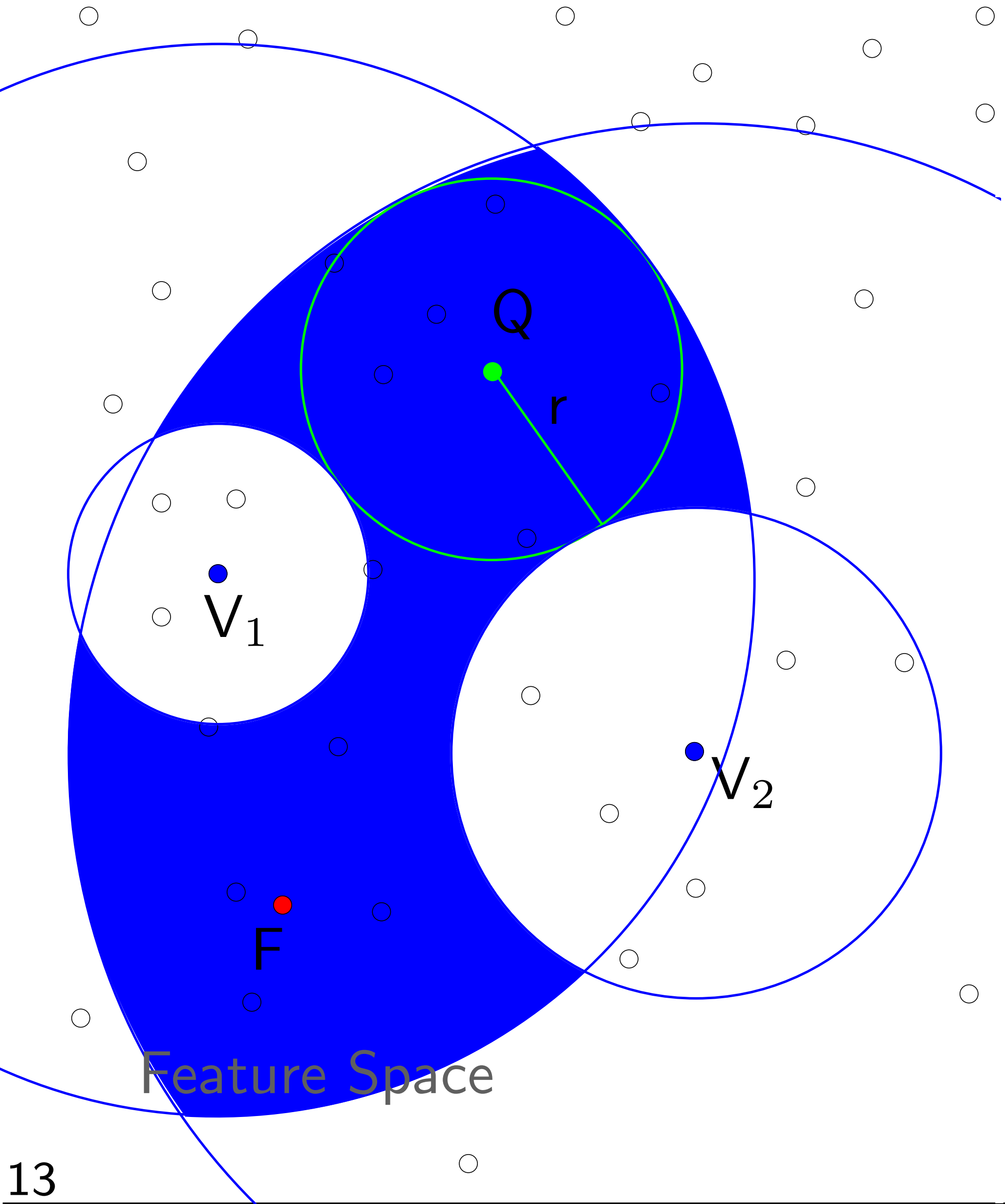
We want to retrieve all objects with a distance $<=$ r from the query object Q.



V_1, V_2: Vantage objects.

Objects with the same coordinates as Q in the Euclidean space lie on the intersections of the black circles around V_1 and V_2.
With two vantage objects, there can be an object $F \neq Q$ with the same coordinates as Q.

We want to retrieve all objects with a distance $<=$ r from the query object Q.



V_1, V_2: Vantage objects.

Objects with a distance $<=$ r from Q in the Euclidean space lie within the intersection of the blue bands around V_1 and V_2.

Search steps:

- Find all objects with an Euclidean distance $<=$ r from Q in the Euclidean vantage space.
- Only for those, calculate the distances to Q in the feature space and discard those with a distance $>$ r.

Future Goals

- **Query-by-Humming without Note Onset Detection:**
  - Represent every note in the database with a point set instead of a single point.
  - Match the sequence of fundamental frequencies from FFT windows with the scores in the database.

  This would avoid the notoriously difficult and error-prone note onset detection.

- **Partial Matching/Tempo Tracking:** Split the queries and the scores in the database into overlapping, polyphonic chunks with a duration of just a few notes and then search for sequences of similar chunks.
  - Tempo variations are possible without requiring explicit tempo tracking. Neither tempo nor measure structure need to be known.
  - The indexing method with vantage objects can be used to make the search for matching chunks efficient.

The Earth Mover's Distance

Measures the minimum amount of work needed to transform one weighted point set into the other by moving weight.

$A = \{a_1, a_2, .., a_m\}, B = \{b_1, b_2, .., b_n\}$: weighted point sets
$w_i, u_j \in \mathbb{R}^+ \cup \{0\}$: their weights.
W=$\sum_{i=1}^{m} w_i$, U=$\sum_{i=1}^{n} u_i$.
f_{ij}: the flow of weight from a_i to b_j over the ground distance d_{ij}.

The set of all possible flows $\mathcal{F} = [f_{ij}]$ is defined by these constraints:

1. $f_{ij} \geq 0, i = 1, ..., m, j = 1, ..., n$ (no negative flow component.)
2. $\sum_{j=1}^{n} f_{ij} \leq w_i, i = 1, ..., m$
3. $\sum_{i=1}^{m} f_{ij} \leq u_j, j = 1, ..., n$

(no point emits or receives more weight than it has.)

4. $\sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} = \min(W, U)$ (the lighter of the two point sets is completely matched.)

$$\mathrm{EMD}(A, B) = \frac{\min_{F \in \mathcal{F}} \sum_{i=1}^{m} \sum_{j=1}^{n} f_{ij} d_{ij}}{\min(W, U)}$$