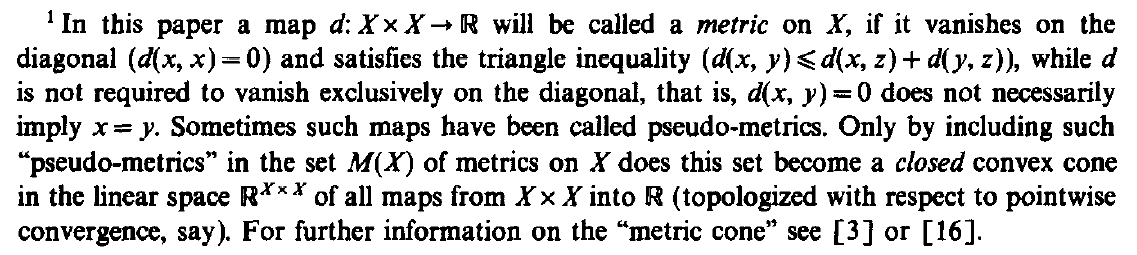
A canonical decomposition theory  
 for metrics on a finite set



Let be a set. Non serve finito per ora.

**Def**. (pseudo-metric)

A function is a **pseudo-metric** on if

that is, it vanishes on the diagonal and it satisfies the triangle inequality.

**Def**. (metric)

A function is a **metric** on if

In particular, a metric is a pseudo-metric.

[BD92a] non parla di non-negatività e simmetria  
 (si possono ricavare da disuguaglianza triangolare + diagonale nulla).

**Prop**. If is a pseudo-metric, then

that is, it is a symmetric and non-negative function.

**Dim**. From triangle inequality on and on

that implies .

From triangle inequality on

that implies .

Let be a vector space over .

**Def**. (convex set)

A subset is **convex** if

**Def**. ((linear) cone)

A subset is a **(linear) cone** if

**Not**.

**Fact** is a (real) vector space.

**Prop**. is a convex cone.

**Dim**. Let pseudo-metric on and . Then

So is a pseudo-metric, , and is a cone.

To show that is convex,  
 we need to show that

But since is a cone,

so it suffice to show that is closed under addition.

Let . Then

So is a pseudo-metric on .

The set of all metrics is also a non-pointed convex cone (same definition of cone but with ): in fact, the zero function is not a metric.

Let us consider with the topology of pointwise convergence .

**Prop**. If is countable[[1]](#footnote-1)\*, then is closed in .

**Dim**. Let us show that is sequentially closed.

Let us consider a convergent sequence of pseudo-metrics

and let us show that the limit is also a pseudo-metric,

For the characterization of the pointwise convergence,

Now for every ,  
 is a real-valued sequence, so

* , so in the limit

This shows that is a pseudo-metric  
 and is sequentially closed.

Since is countable (and so is ), the space   
 is a countable product of first-countable spaces (namely ), so it is first-countable itself.

Now is sequentially closed in a first-countable space,  
 so it is closed.

If is uncountable, then we cannot conclude that is first-countable.

**Oss**. In the same hypothesis, the set of metrics is not even sequentially closed in .

**Dim**. Let us consider a convergent sequence of metrics such that

for some . Then

but , so is not a metric.

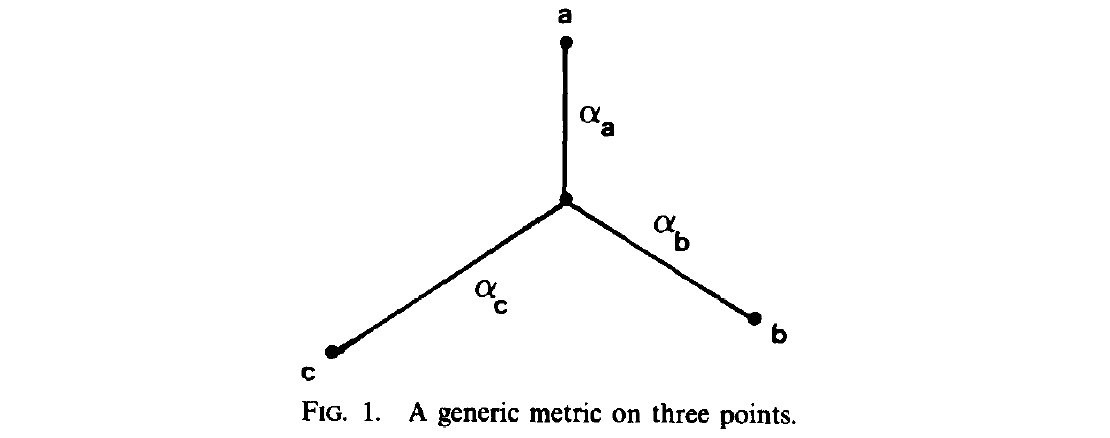
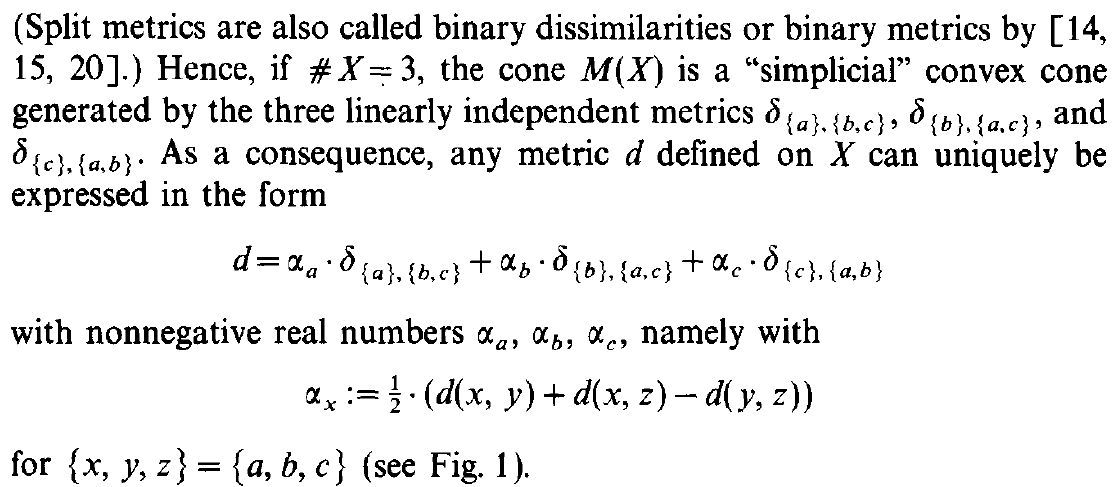
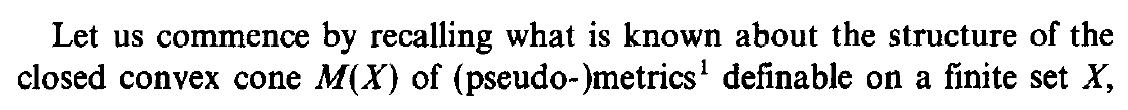
<https://en.wikipedia.org/wiki/Metric_space>  
<https://en.wikipedia.org/wiki/Pseudometric_space>

<https://en.wikipedia.org/wiki/Vector_space>  
<https://en.wikipedia.org/wiki/Convex_set>  
<https://en.wikipedia.org/wiki/Convex_cone>

<https://en.wikipedia.org/wiki/Pointwise_convergence>  
<https://en.wikipedia.org/wiki/Sequential_space>

Sometimes it is useful to think of functions   
 as real-valued square matrices .  
In particular, a pseudo-metric corresponds to a symmetric matrix,  
 with 0 on the diagonal, non-negative entries elsewhere  
 and such that the elements satisfy the triangle inequality.

Chapter 1: Introduction



Let us consider a real vector space.

**Def**. (conic/conical combination)

A point is a **conical combination** of if

such that

**Def**. (extreme/extremal ray)

An **extreme ray** of a cone is a subset of the form

for some , such that the elements of cannot be expressed as conical combinations of elements of ;

or equivalently, for every and

or equivalently, for every

We can identify an extreme ray with the associated vector .

**Def**. (simplicial cone)

A cone is a **simplicial cone** if every complete[[2]](#footnote-2)\* set of representatives of the extreme rays is linearly independent;

or equivalently, if

where is the vector subspace spanned by .

From now on we will assume finite of cardinality .  
We may sometimes identify with , since they are in bijection.

**Def**. (split metric / binary metric / binary dissimilarity)

Given a partition (or split) of into two disjoint non-empty sets and , we call **split metric** of the function defined by

**Not**. If , we denote the trivial split metric

**Prop**. The split metrics are pseudo-metrics.

**Dim**. Let us fix a split of and let us show that . The vanishing on the diagonal is obvious.

Let us show the triangle inequality.



Analogous cases switching and .

**Oss**. Given , for every we have the following triangle inequalities

If , from the last two we get .

**Not**. For every split of , we define

These are the pseudo-metrics that vanish  
 where the split metric vanishes (the other entries can be anything).

**Lemma** These pseudo-metrics are multiples of the relative split metric

**Dim**. Let us fix .  
Since , we have

Symmetrically, for we have

Thus .

This means that either vanishes everywhere or, where does not vanish, it can assume only one value – say ; and so

**Prop**. The split metrics are extreme rays of .[[3]](#footnote-3)\*

**Dim**. Let us consider a split and the associated  
 split metric . From the previous **Lemma** we have

so we need to show

Suppose . Then for every

but since are non-negative, we have

Idem for every .

Thus and vanish where vanishes, that is

**Oss**. The number of split metrics is .

In fact, the set of split metrics is in bijection with the set of splits of . In creating a split, for every element of we have two choices: put it in or put it in .  
We have possible arrangements. We need to subtract the cases corresponding to and , and divide by two, since the split is the same as .   
In the end we get

**Lemma** The functions defined by

can be expressed as linear combinations of the split metrics

**Prop**. is a vector subspace of with the following properties:

**Dim**. Let us indicate the set of symmetric functions vanishing on the diagonal with .   
It is clear that and is closed under linear combinations. So .

Notice that is a basis for . In fact, the functions in can be represented as symmetric matrices with zeroes on the diagonal; while the function can be represented as a symmetric matrix with in positions and , and zeroes elsewhere.

This also shows that

From the previous **Lemma** we can express the functions as linear combinations of split metrics, thus every function in ; that is .

**Oss**. For , say , all the pseudo-metrics  
 are multiple of the only split metric ;   
so is just a one-dimensional ray.

For , say , the split metrics

generate the cone .  
In fact, let us consider a pseudo-metric ;  
then we want to show that such that

In particular, by evaluating

and analogously for the other couples,   
we get the following system of equations

Solving for we get

Moreover, from triangle inequality on , we have .  
This shows that these split metrics are the only extreme rays (every other pseudo-metric is a conical combination of them).

The same calculation also shows that  
 the split metrics are linearly independent in . In fact, if

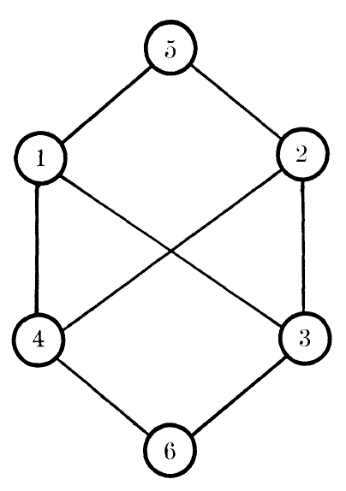
where is the identically zero pseudo-metric, then

In particular, for the decomposition in split metrics is unique.

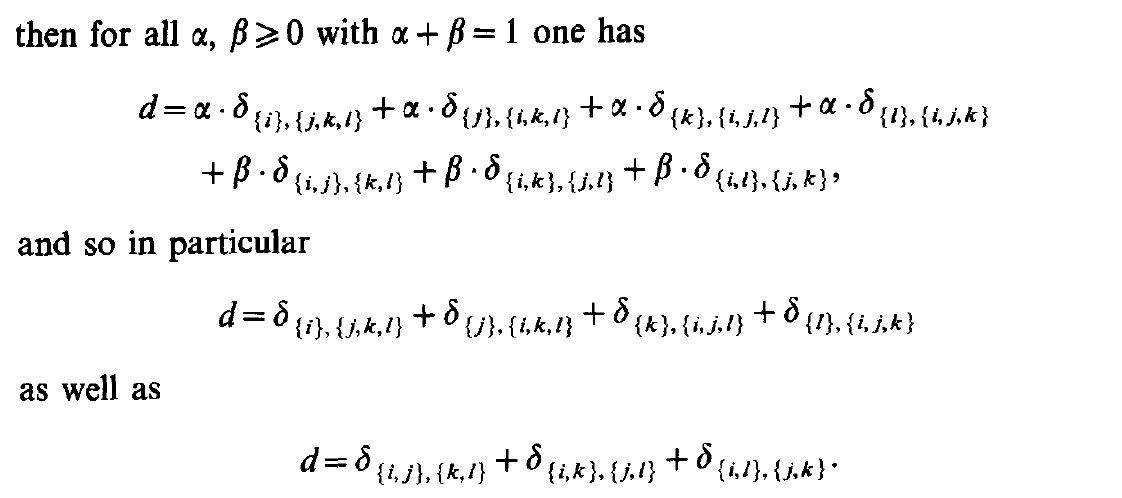
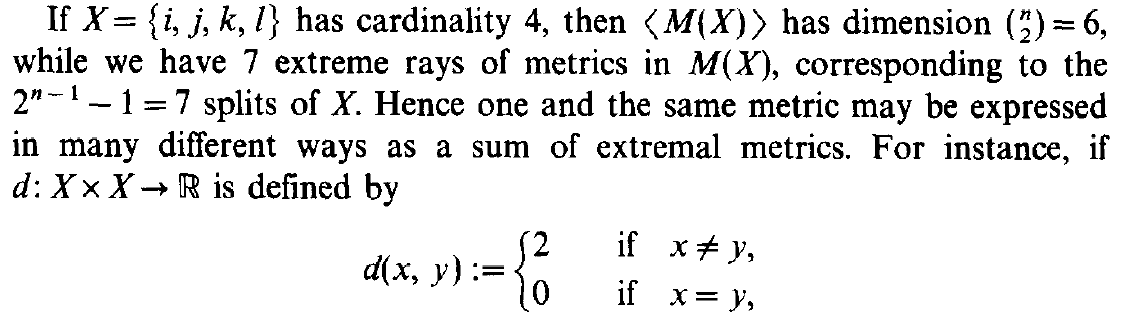
**Oss**. is a simplicial cone if and only if . In fact,

and for this number is greater than

But for the number of extreme rays coincide with the number of split metrics and

Esistono metriche estremali che non sono split? L’esempio a p. 17 di [Avis80] sembra mostrare nel caso una metrica (indotta da un grafo) che è estremale ma non split (il grafo è bipartito ma non completo).

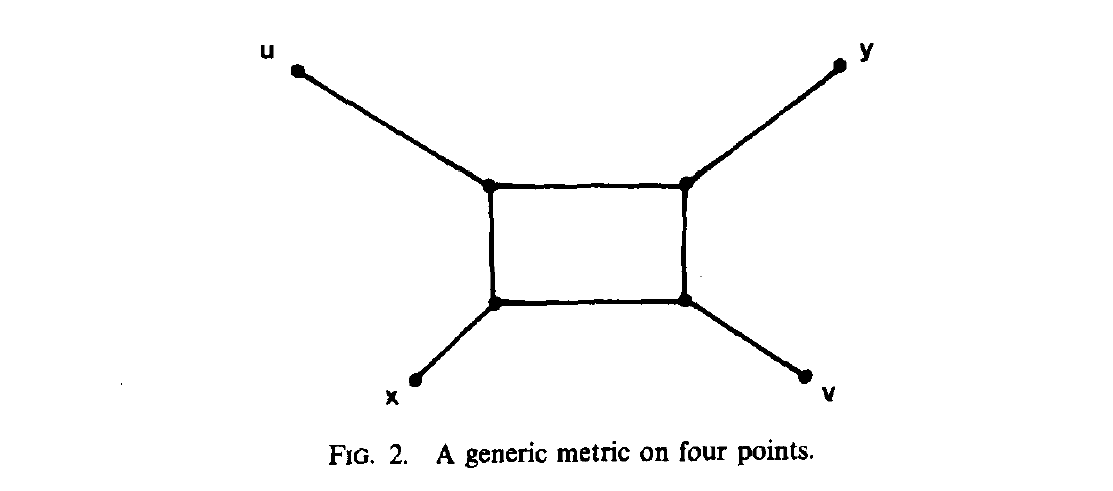
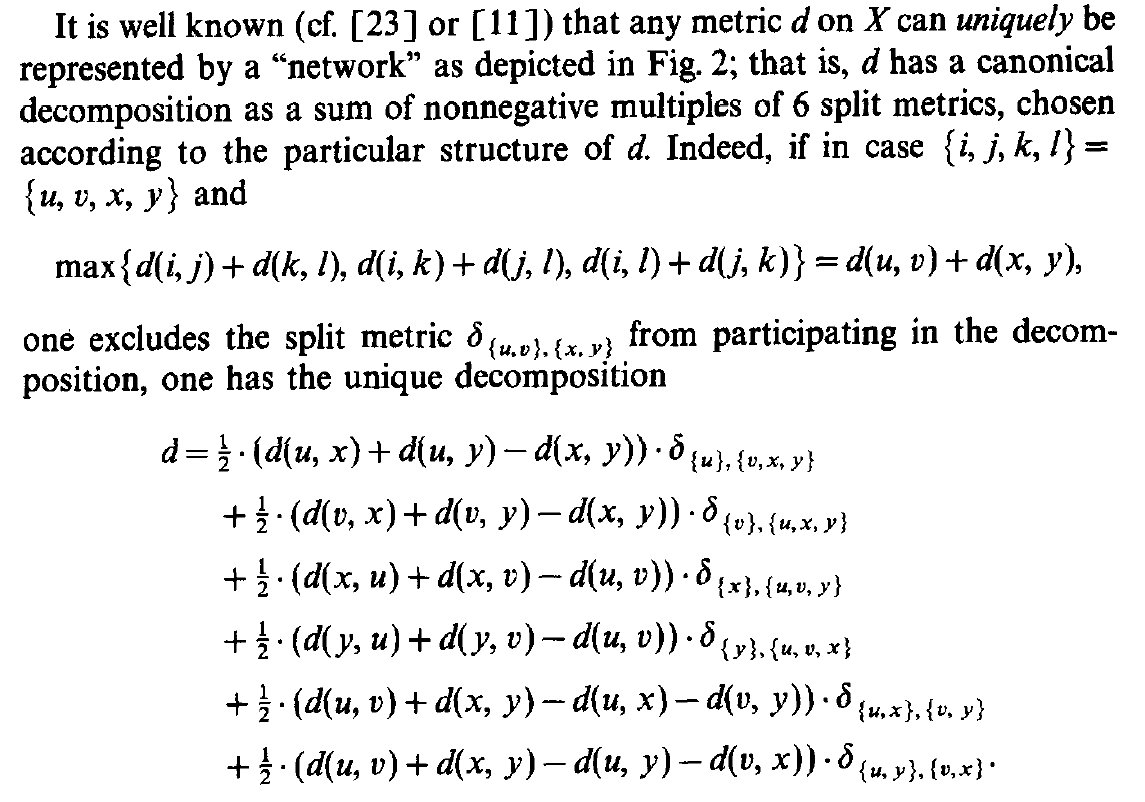
Infatti le split metric si possono rappresentare come grafi bipartiti completi.



The counterexample above shows that for the decomposition in split metrics is not necessarily unique.

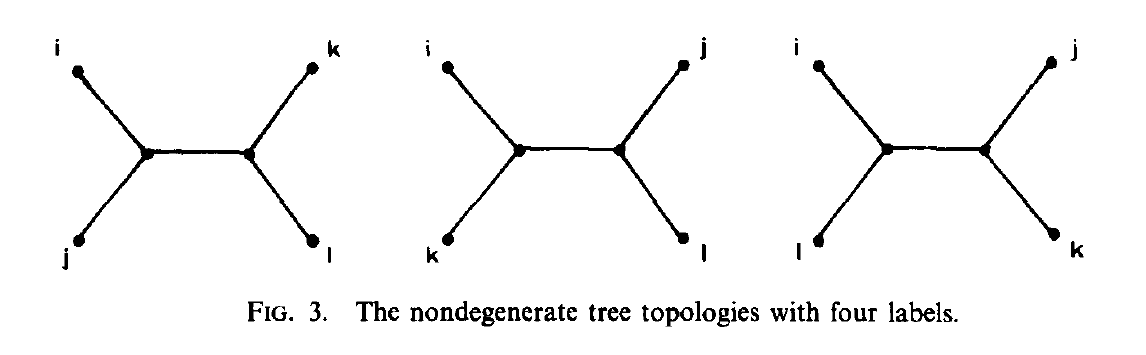
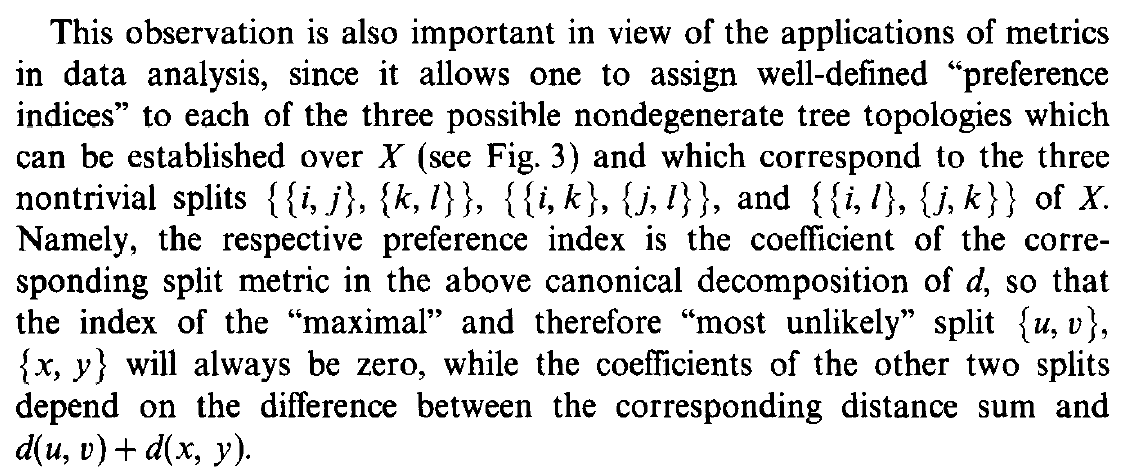
<https://en.wikipedia.org/wiki/Conical_combination>  
<https://www.damtp.cam.ac.uk/user/hf323/M18-OPT/lecture2.pdf>  
<https://people.math.carleton.ca/~kcheung/math/notes/MATH5801/10/10_2_extreme_rays.html>

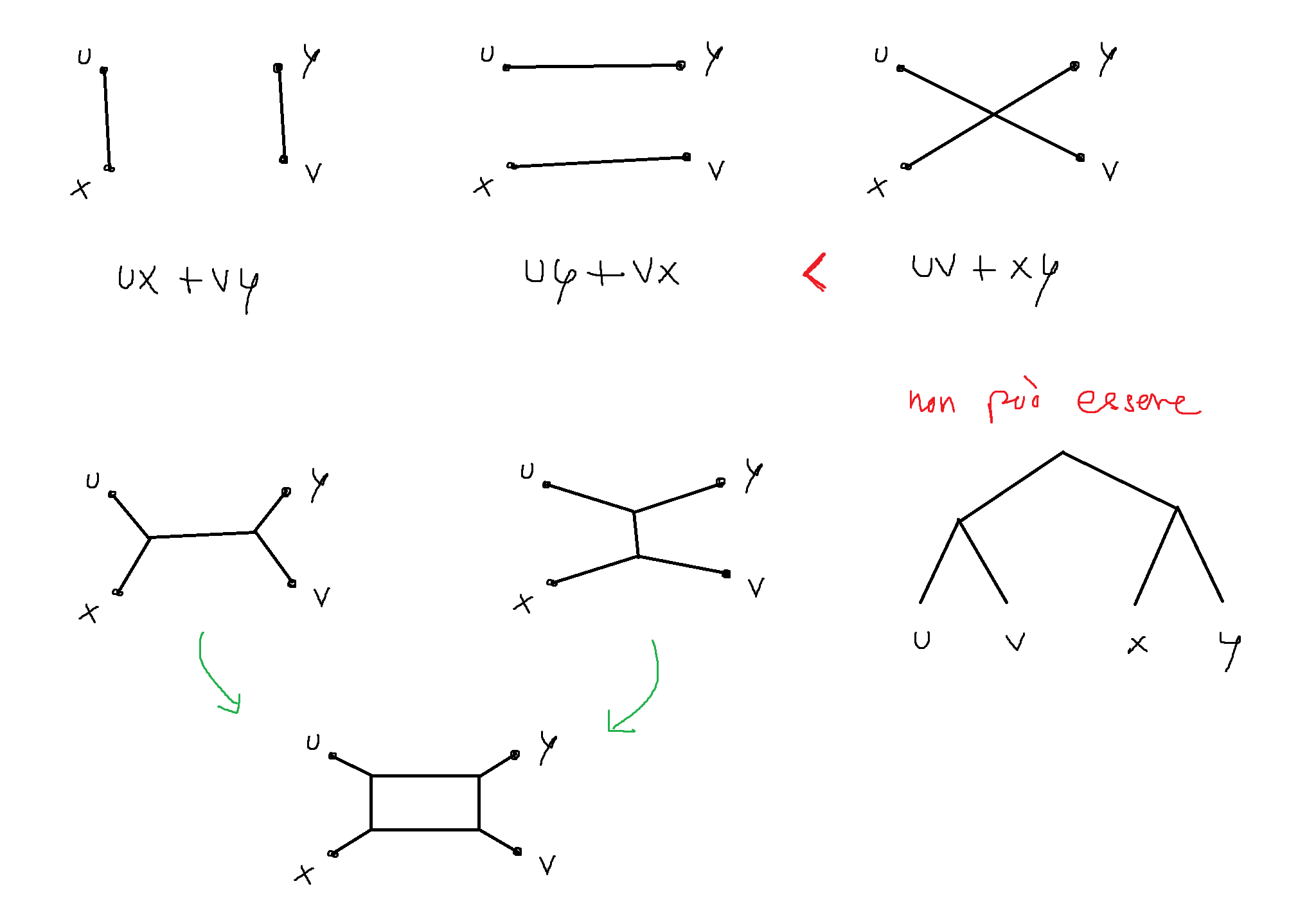
<https://www.mat.uniroma2.it/~tvmsscho/Rome-Moscow_School/2012/files/RM12-Cones-Matrices.pdf>  
<https://math.stackexchange.com/questions/4362253/simplicial-polyhedral-cones>

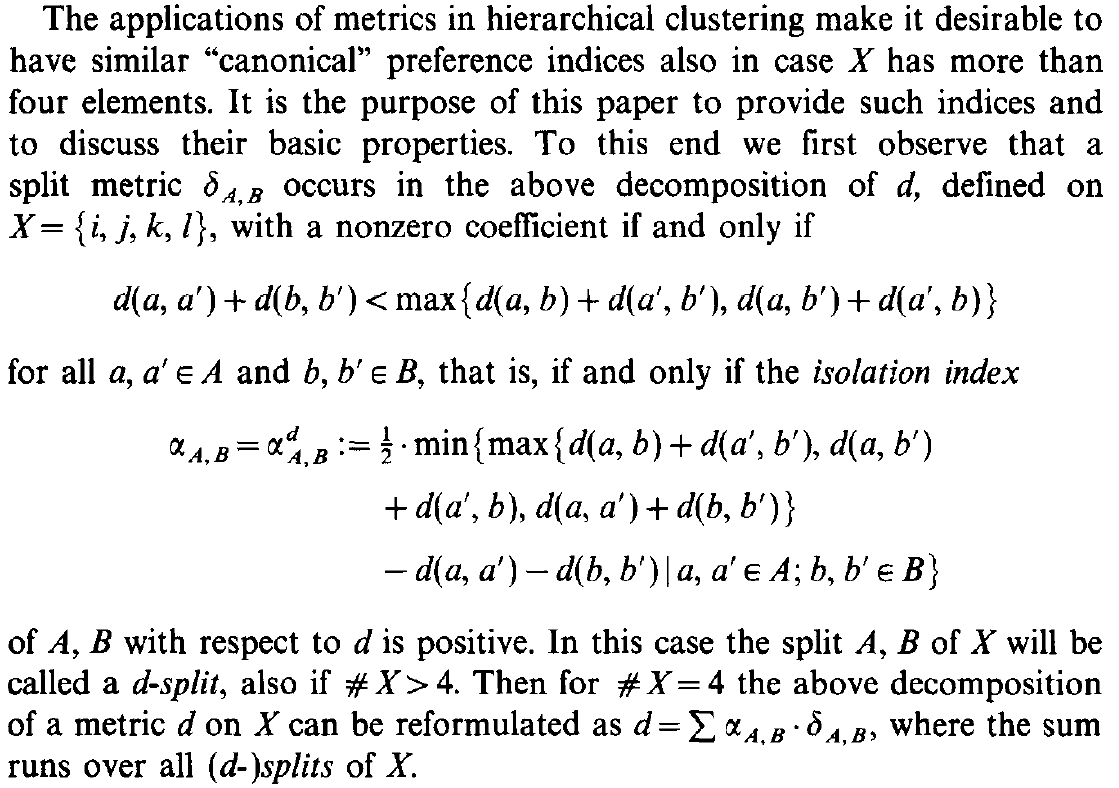


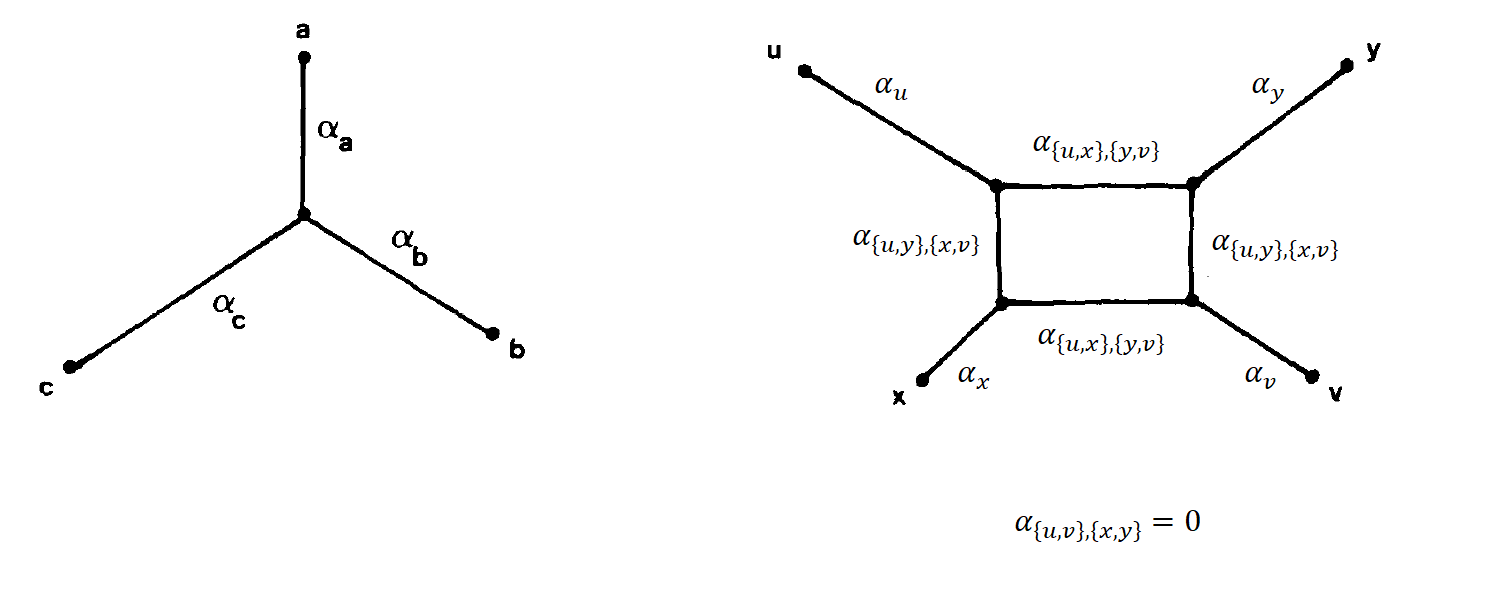
Dimostrare la decomposizione.

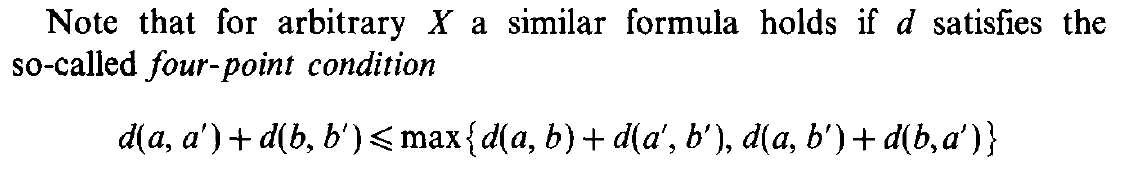
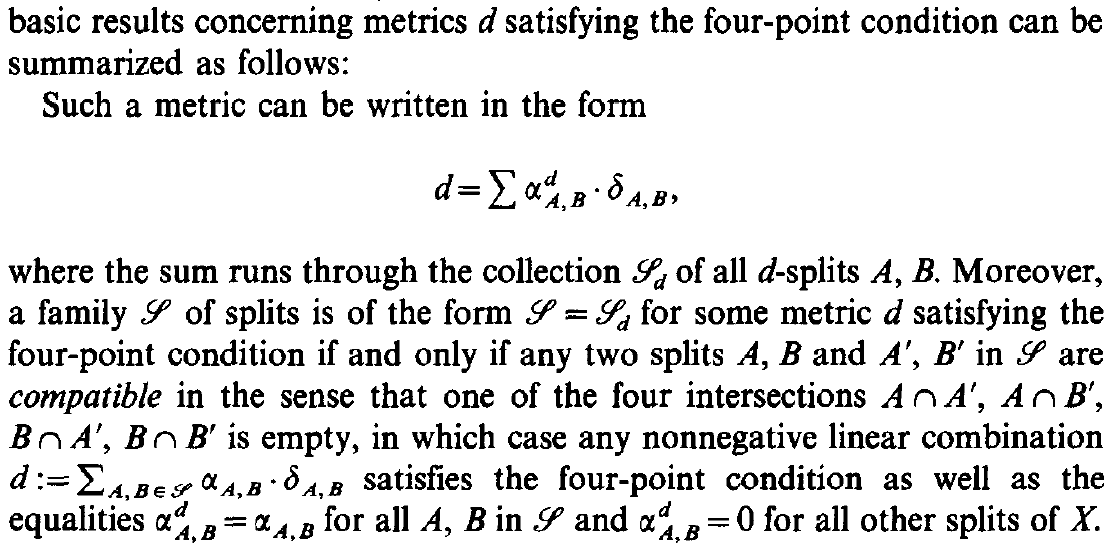
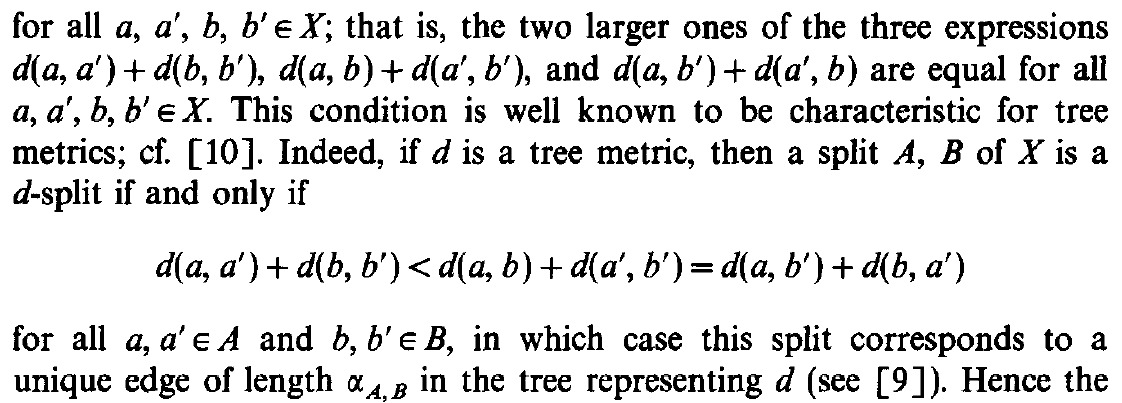
Per spiegare la roba che viene dopo è meglio aver letto il resto del paper.









**Fact** A pseudo-metric is a tree metric if and only if it satisfies  
the four-point condition, that is for every

**Fact** Let be a tree metric. Then a split is a -split if and only if for every and

In this case the split corresponds to a unique edge  
 of length in the tree representing .

**Prop**. If is a pseudo-metric satisfying the four-point condition,  
 then it can be written as

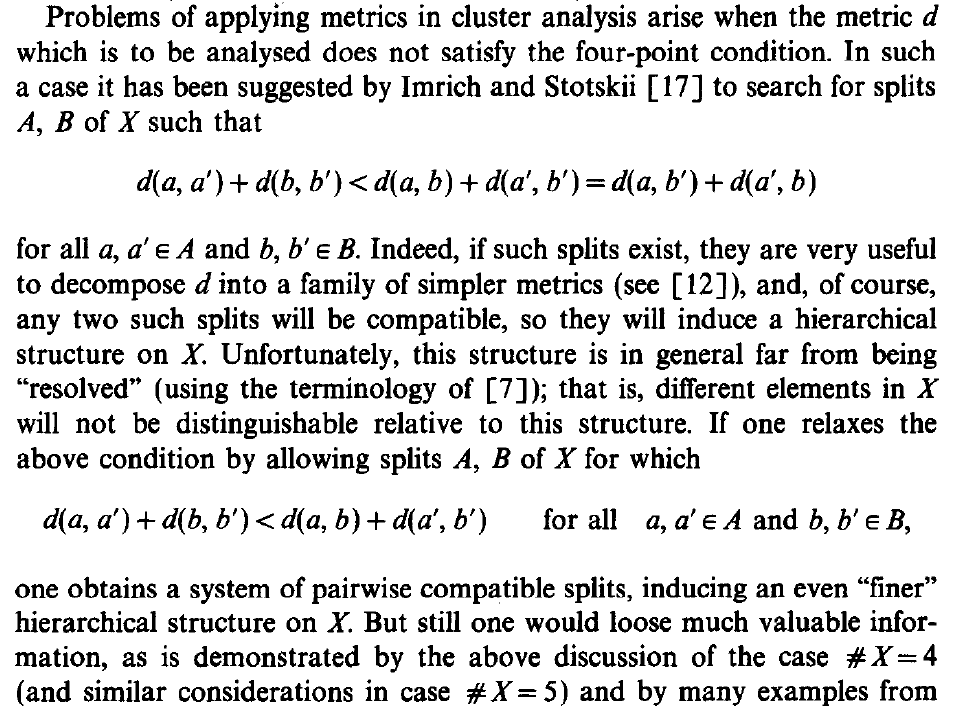
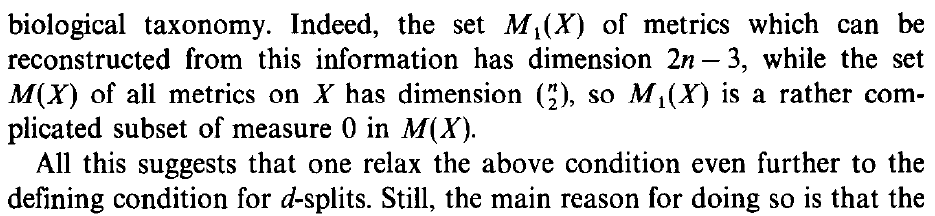
Moreover, a collection of splits is of the form   
 where is a pseudo-metric satisfying the four-point condition, if and only if the splits in are compatible.

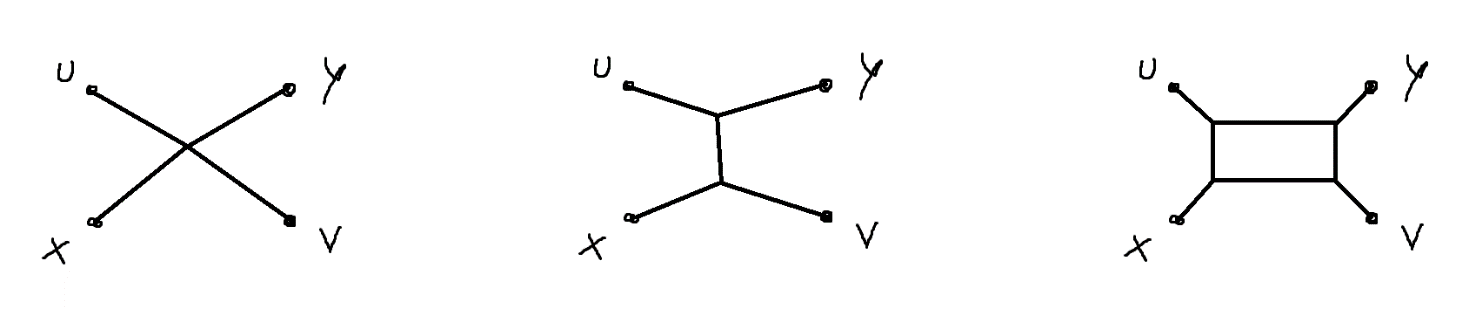
**Dim**. The first assertion is a consequence of **Teo**. 2 and **Cor**. 7.

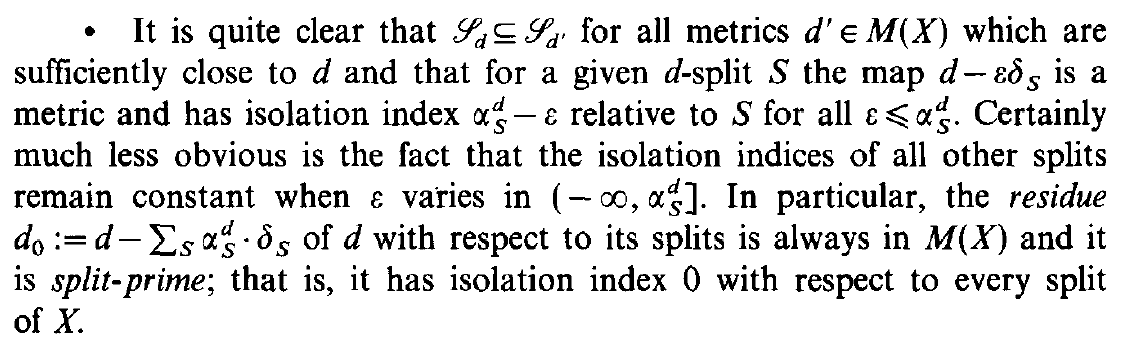
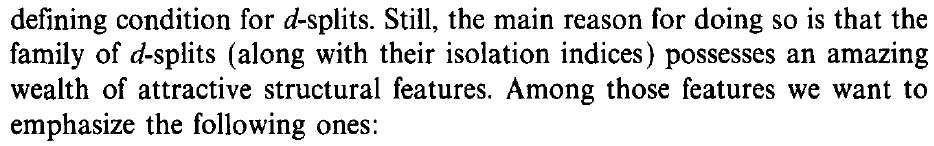
It is a consequence of **Cor**. 7.

Since compatible implies weakly compatible,  
it is a consequence of **Teo**. 3 + **Cor**. 7.

In other words, tree metrics are totally decomposable and  
a set of splits coincide with the -splits of some tree metric if and only if  
 they are compatible.



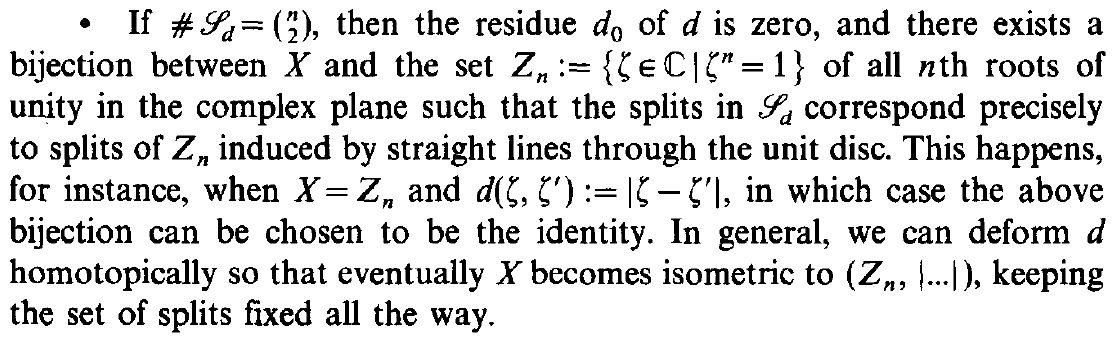


**Prop**. Let be a pseudo-metric, a -split and .

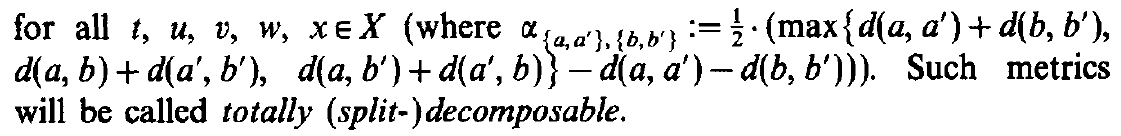
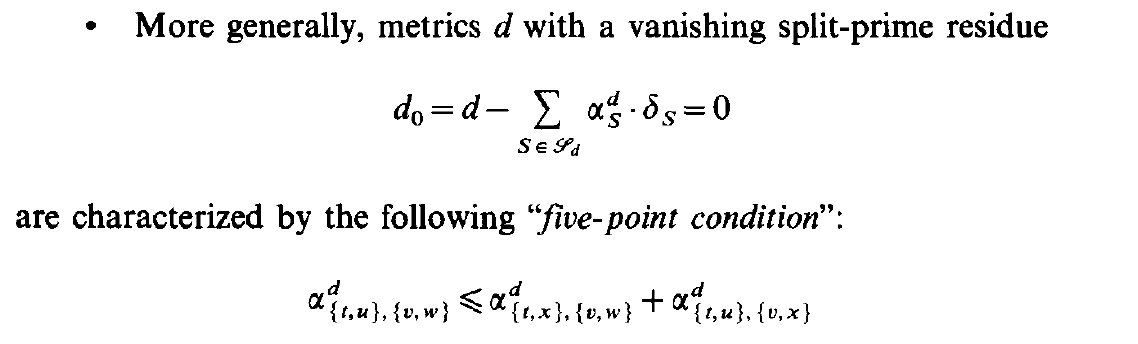
Then is a pseudo-metric,  
 and .

In particular, the residue   
 is a pseudo-metric and it is split-prime,  
 that is .

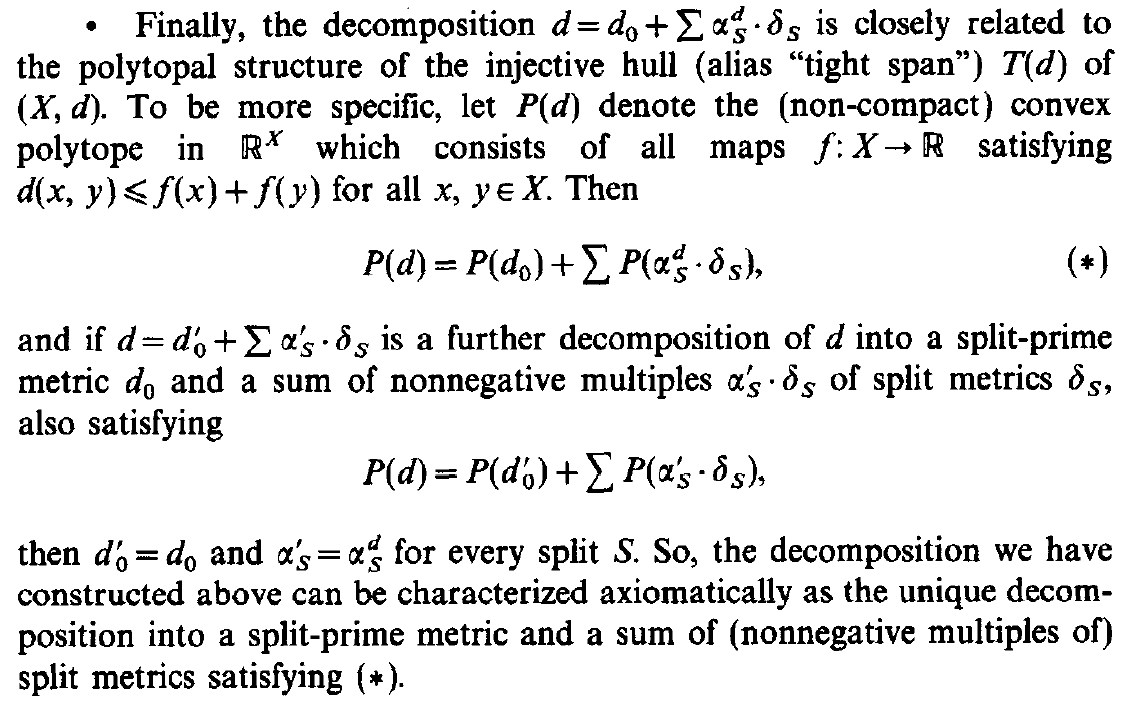
**Dim**. It is a consequence of **Teo**. 2.



La prima riga è il **Cor**. 5; il resto è conseguenza di (ii) -> (i) del **Teo**. 5 con una riformulazione di circular split. L’ultima frase non viene dimostrata.

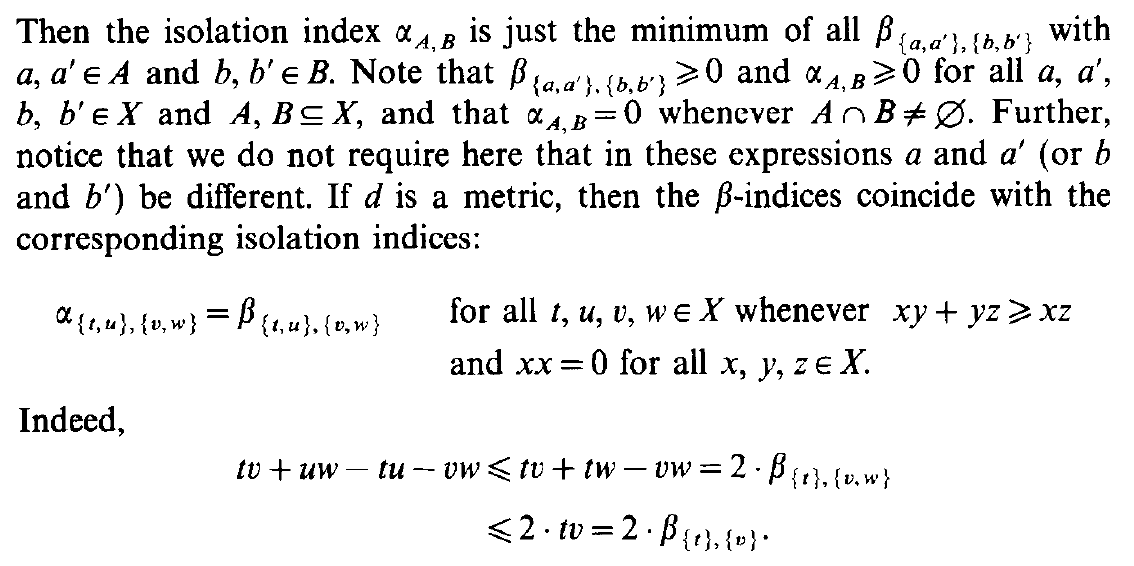
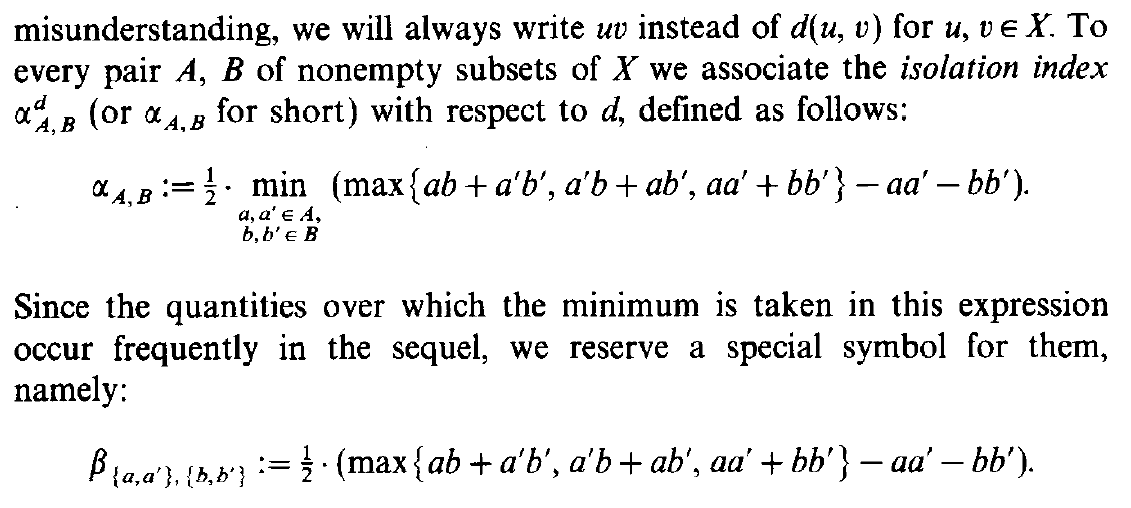
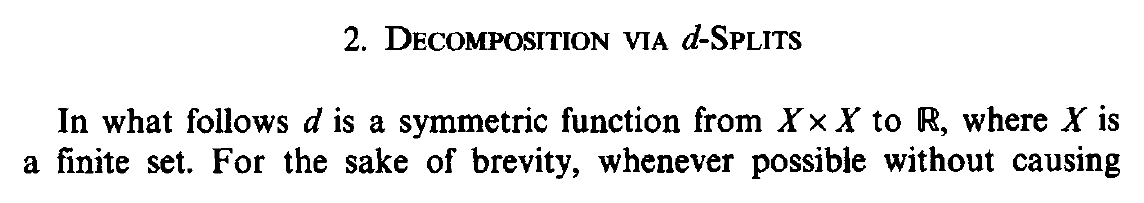


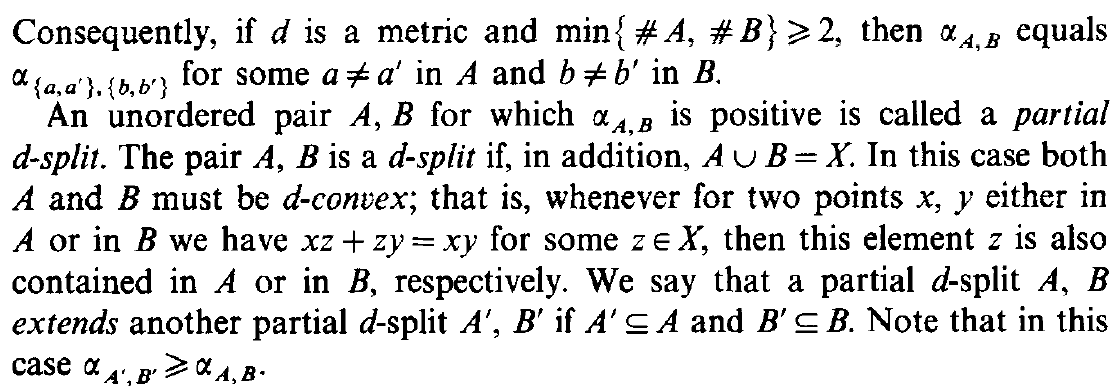
Questo è (i) => (iv) del **Teo**. 6.



Questo è il **Teo**. 8.

Chapter 2: Decomposition via -Splits





Let be a symmetric function.

**Not**.

In the following we will implicitly refer to .

**Def**. (isolation index)

Let be non-empty subset of . Then for every and

We call the **isolation index** with respect to .

**Oss**. We have   
since we are subtracting a term that is inside the maximum:

It follows that also .

**Oss**. If and intersect, then .

In fact, let . Then

and so .

**Prop**. If is a pseudo-metric, then

**Dim**. By (reverse) triangle inequality and the fact that vanishes on the diagonal, we observe that

where we used in the first line  
 and in the second one.

With analogous calculations we get

Again by (reverse) triangle inequality

Since

we have proven that is smaller than all other possible indices for the quartet .

This implies by definition of the isolation index

**Cor**. If is a pseudo-metric and   
 (that is both and have at least 2 elements), then

for some and .

**Dim**. Let be the quartet that minimizes the index.

Then, for the previous inequalities about indices  
 and the hypothesis on the number of elements of and ,   
we conclude that and .

Using the previous **Prop**.

**Def**. (splits and -splits)

A **partial split** (of ) is a pair with .  
A **(total) split** (of ) is partial split such that .

A partial/total **-split** is a partial/total split such that .  
Notice that in this case and must be disjoint.

We call a split **trivial** if one of the parts contains only one element.  
A partial split of the form is called **quartet**.

We denote the set of all splits of with   
 and the set of all -splits of with .

Given and a split of , we have .  
For this reason, we may refer to the -splits as -split of   
 and indicate them with .

**Def**. (-convex set)

A set is **-convex** if

**Prop**. If is a -split, then and are -convex.

**Dim**. Let and such that .

Since , we have or .  
If by absurd , then

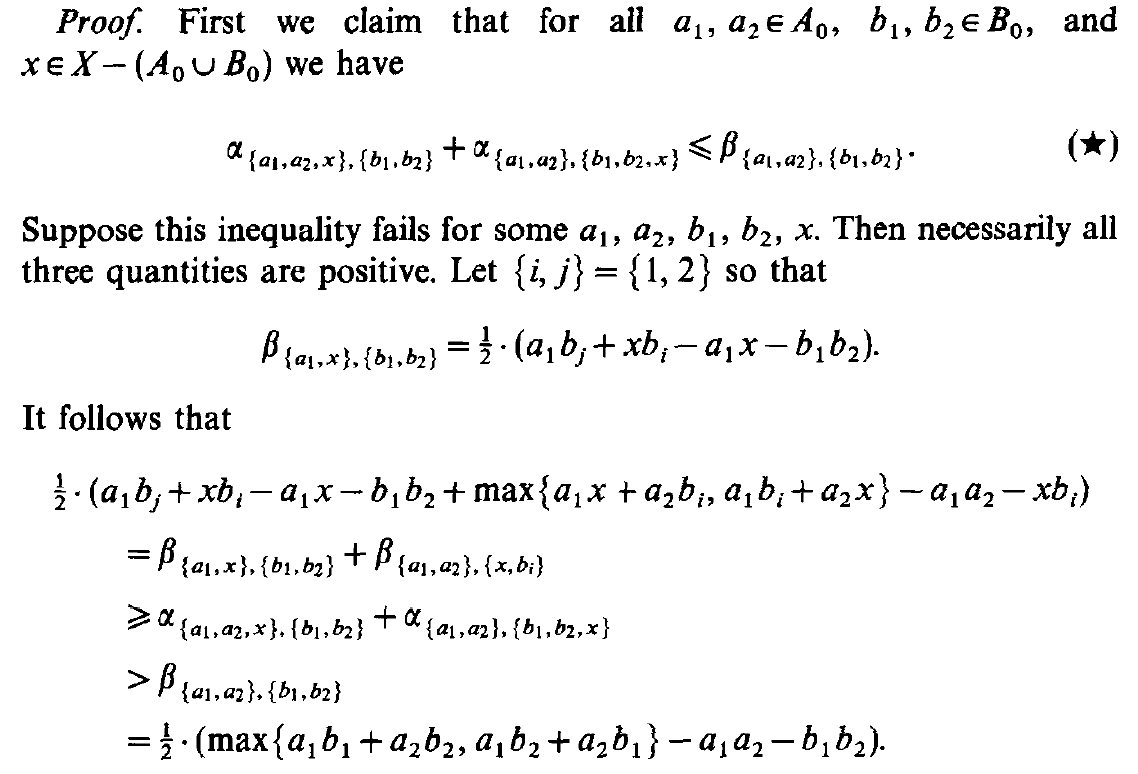
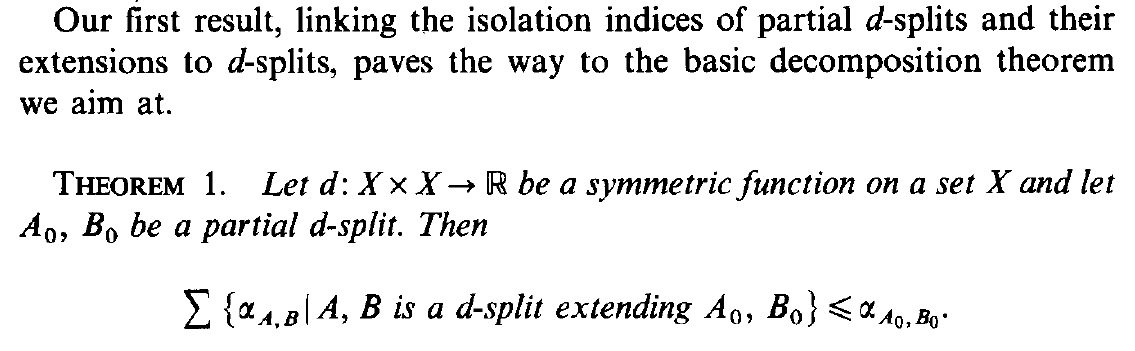
that implies . ⭍

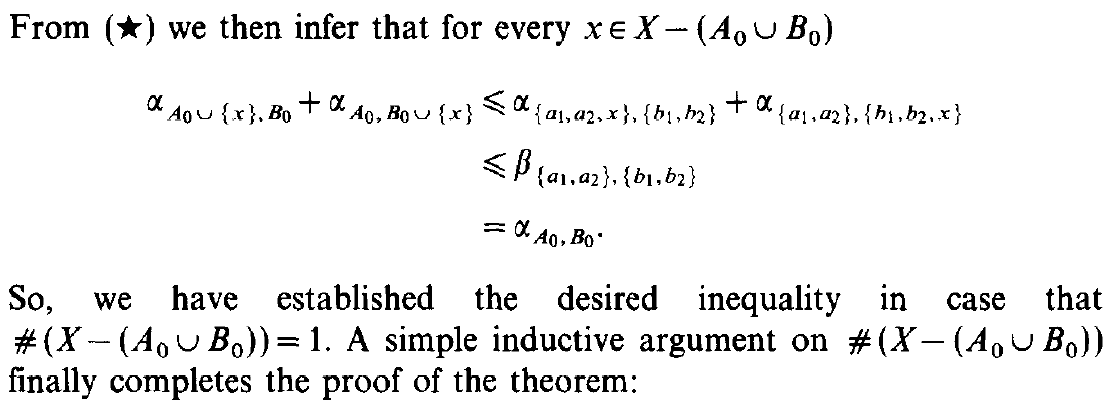
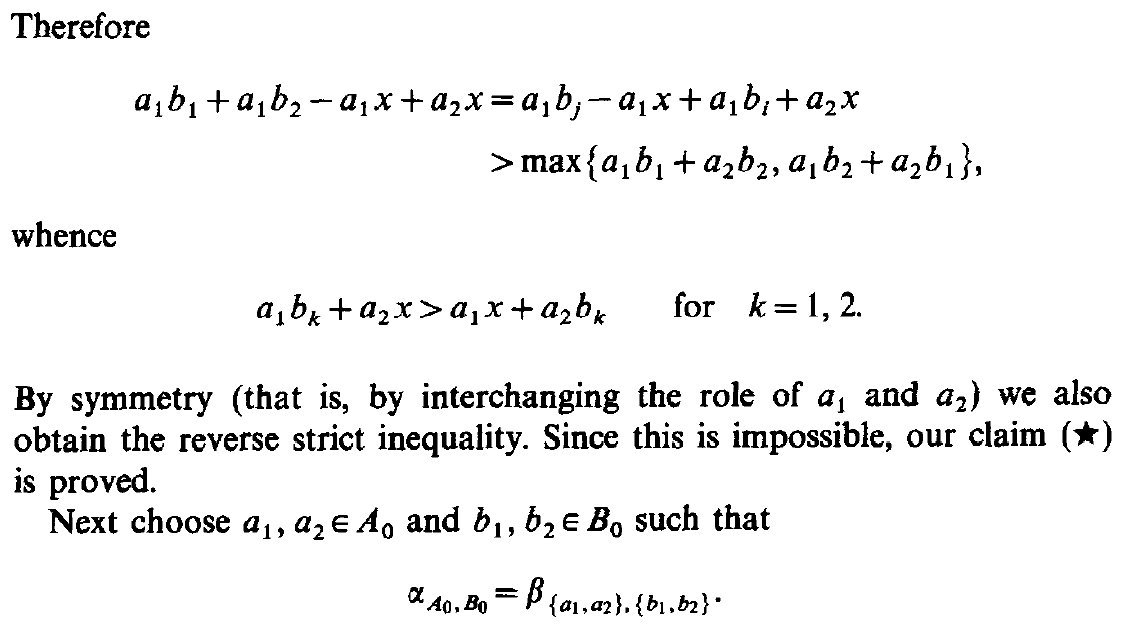
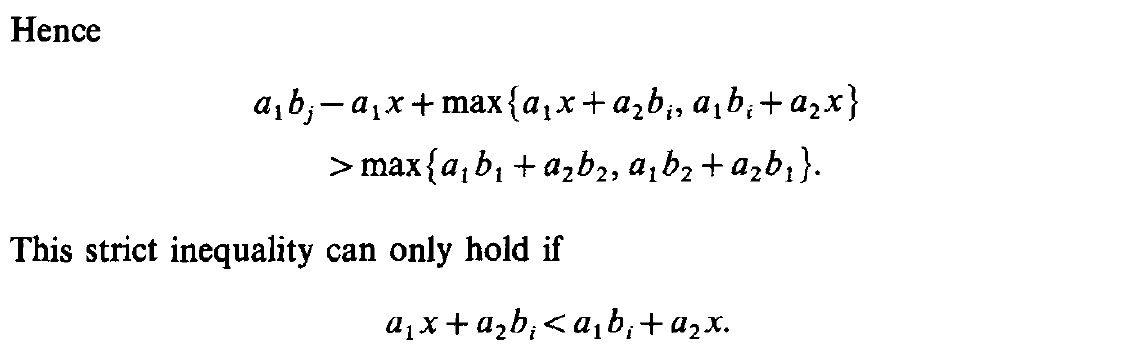
Then it must be and so is -convex.

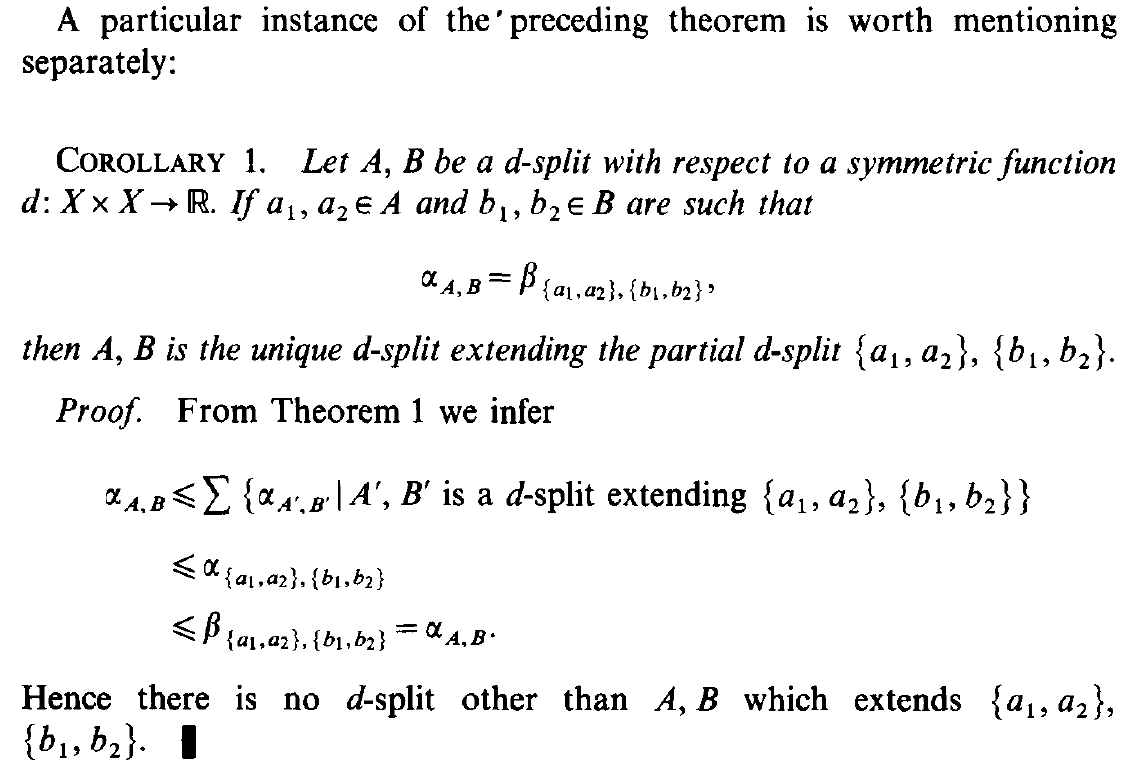
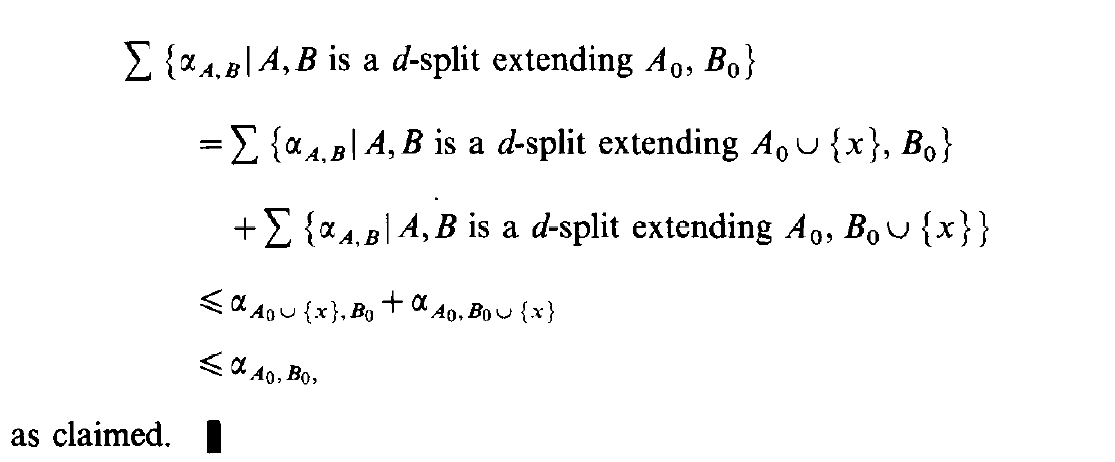
**Def**. (extension of -splits)

We say that a partial -split **extends** another partial -split   
 if and (or and ).   
We denote it .

Notice that (it’s a minimum on a larger set).







**Lemma** Let be a partial -split.   
Then for every

**Dim**. Suppose by absurd that

Observe that since is a -split, so also

From this we have

Let so that

in fact, since , the maximum

cannot be .

For the same reason we have

From the previous inequalities we have

and by substituting the expressions for the indices

simplifying we obtain

If by absurd ,  
the previous inequality would become

but since is a term of the maximum on the right, this is a contradiction. ⭍

So we must have and thus

For this to be true we need

By symmetry, interchanging and ,  
 we obtain the other strict inequality. ⭍

**Teo**. 1 Let be a partial -split. Then

**Dim**. Let and such that

From the previous **Lemma** we have

This proves the theorem in the case .

We prove the general case by induction on .  
We have already seen the base case.

Let and .

Observe that the -splits extending are exactly   
the union of those extending and those extending .   
In fact, since -splits are partitions, if extends (say WLOG and ), it must be or ;  
that is or .

Moreover,

So we can apply the inductive step:

**Oss**. We can substitute splits with -splits in the sum  
 (splits that are not -splits do not contribute).

**Cor**. 1 Let a -split and let and such that

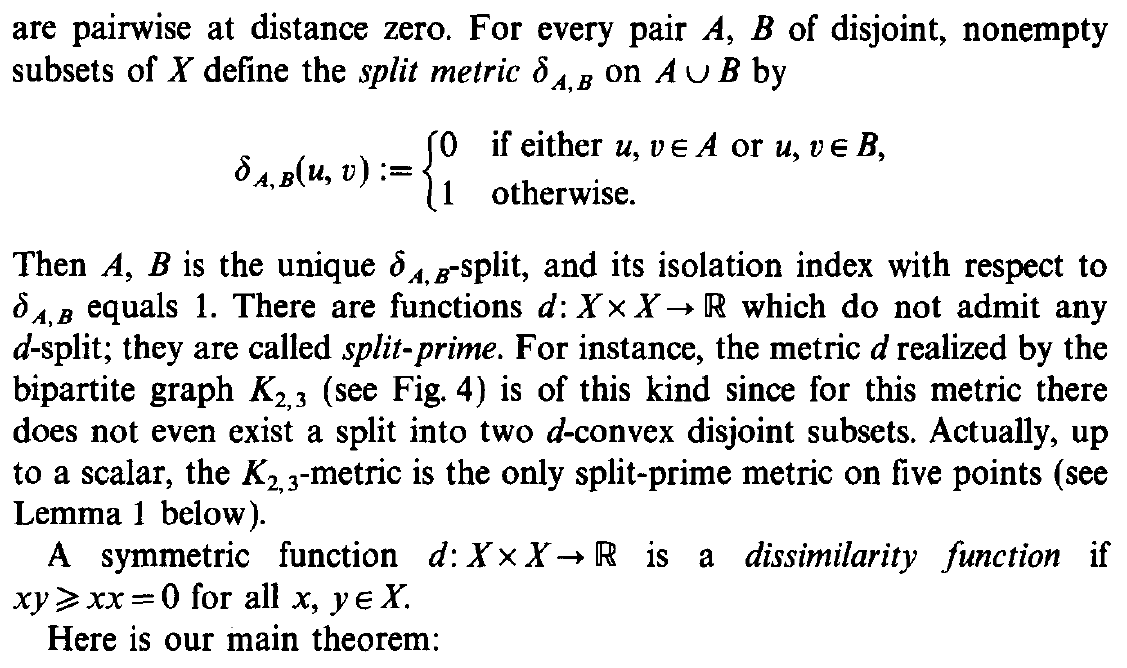
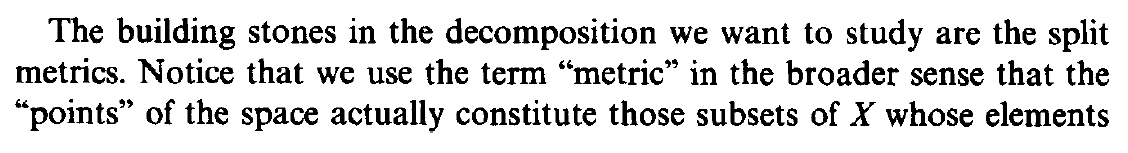
Then is the unique -split that extends .

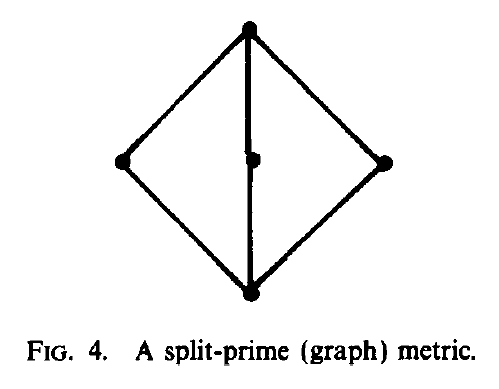
**Dim**. Observe that is a partial -split.  
In fact, if by absurd

then , but is a -split. ⭍

By applying **Teo**. 1

Since the sum is equal to and is a member of the range of the sum, this means that is the only -split that extends .





**Def**. Given non-empty disjoint subsets of ,   
the **split metric** on is defined as

Questa definizione è leggermente diversa da quella data precedentemente perché non si richiede .

**Oss**. is the only -split and its isolation index is 1.

In fact, let us consider and another split , with the following intersections: .  
Since and ,  
at least two of these intersection must be non-empty.   
We can divide in cases.

If all the intersections are non-empty, we can pick

then, with respect to , we have

so .

If only three of the intersections are non-empty  
 – say WLOG – then we can pick

then, with respect to , we have

so .

If only two intersections are non-empty  
 – say WLOG – then .   
In fact, since and ,  
 then we have and .  
But since , it must be and .

Now for every and (with respect to )

so .

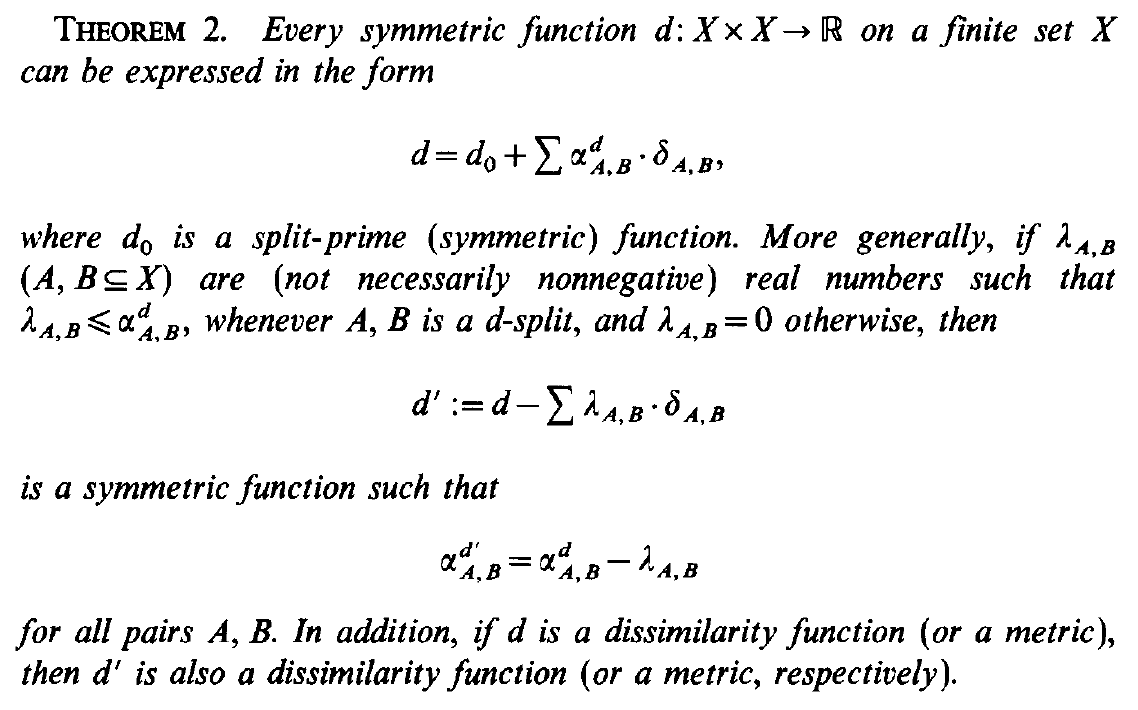
**Def**. (split-prime function)

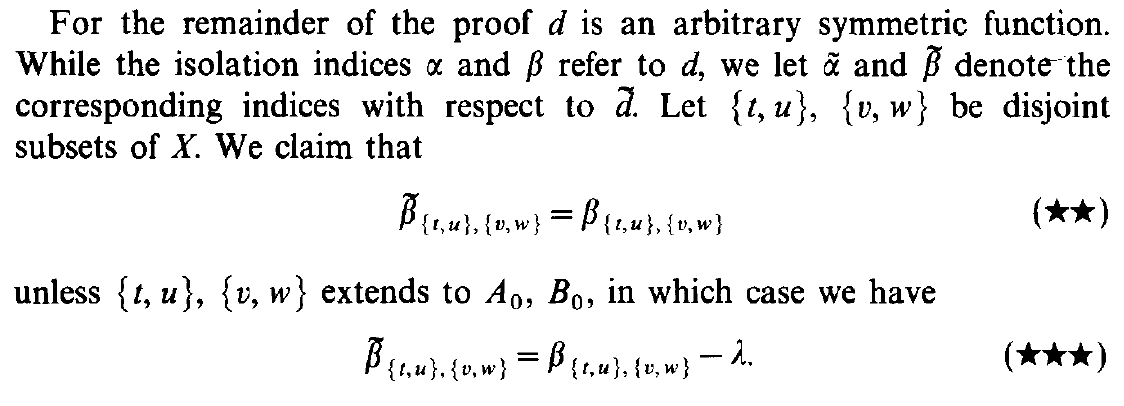
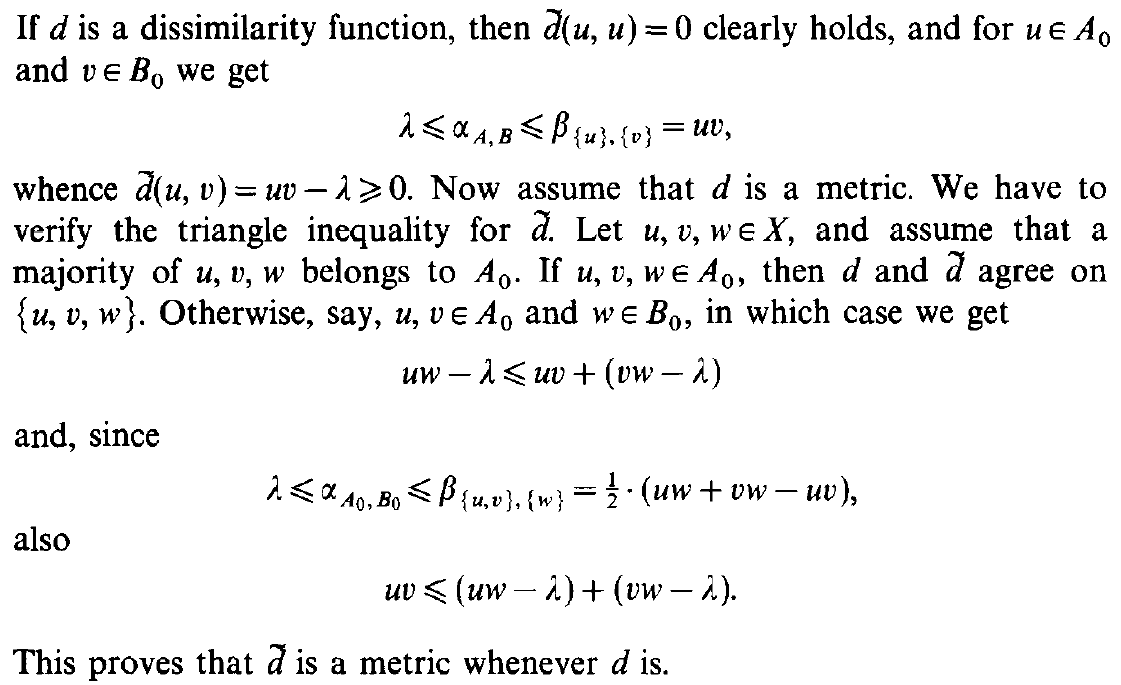
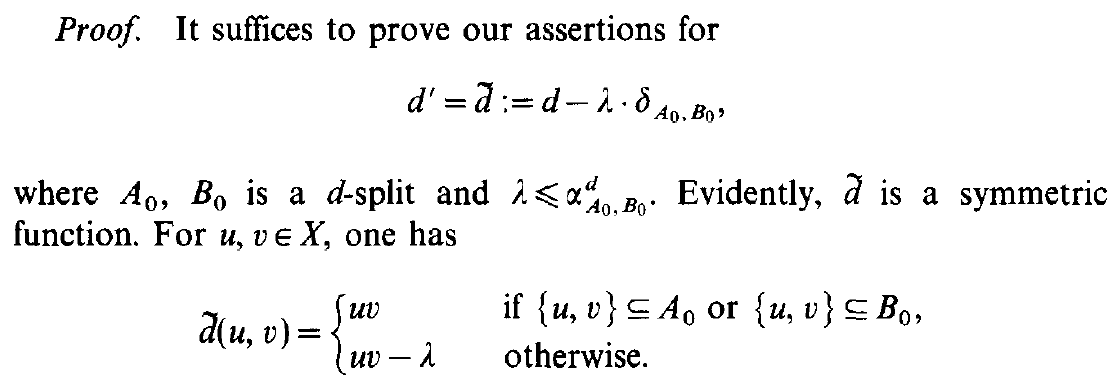
A function is **split-prime** if it does not admit any -split.

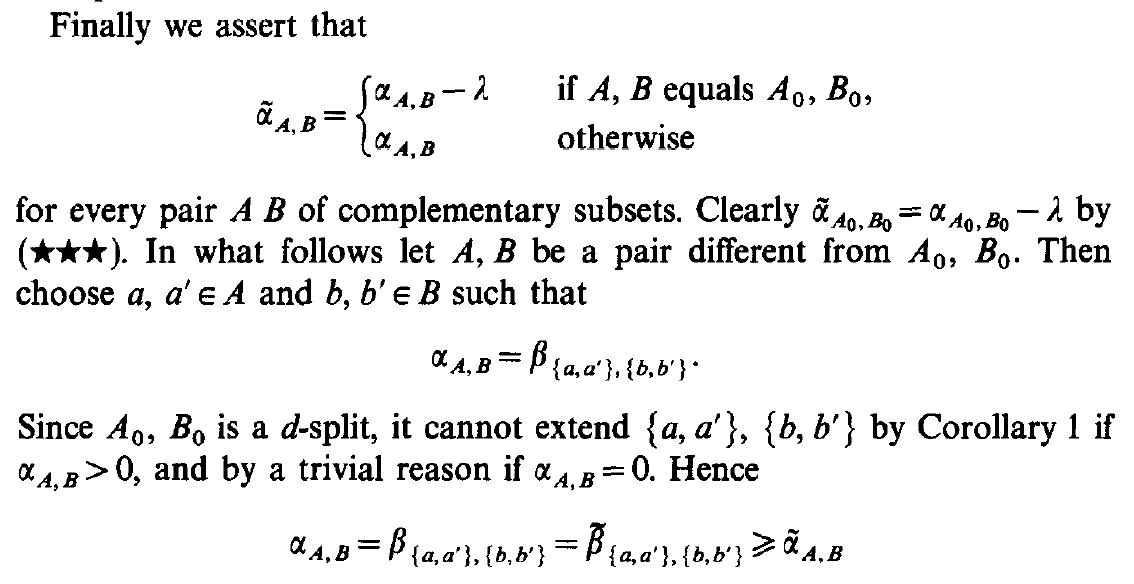
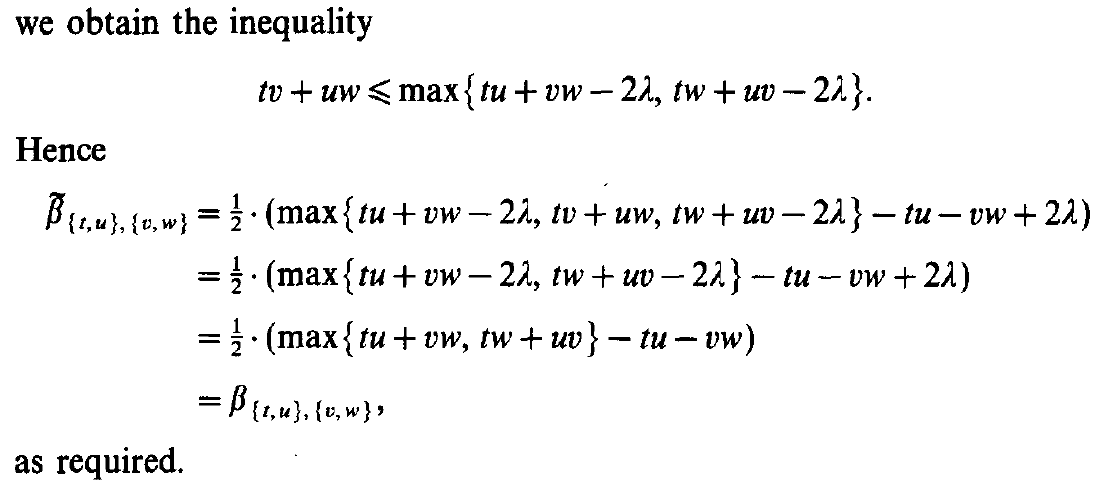
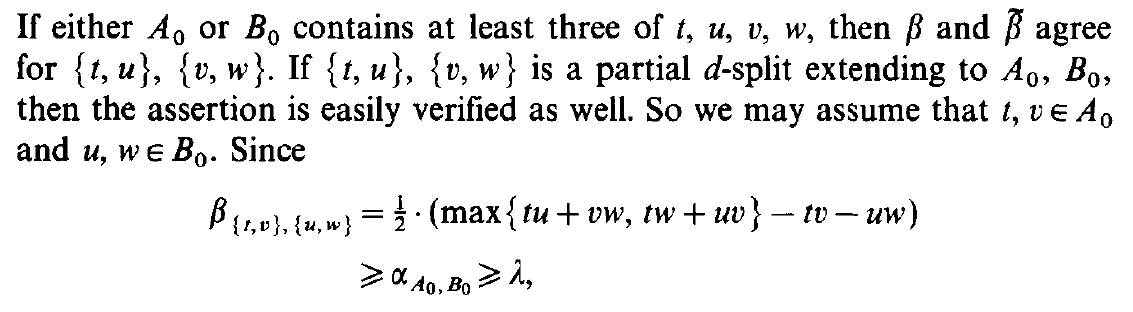
**Def**. (dissimilarity function)

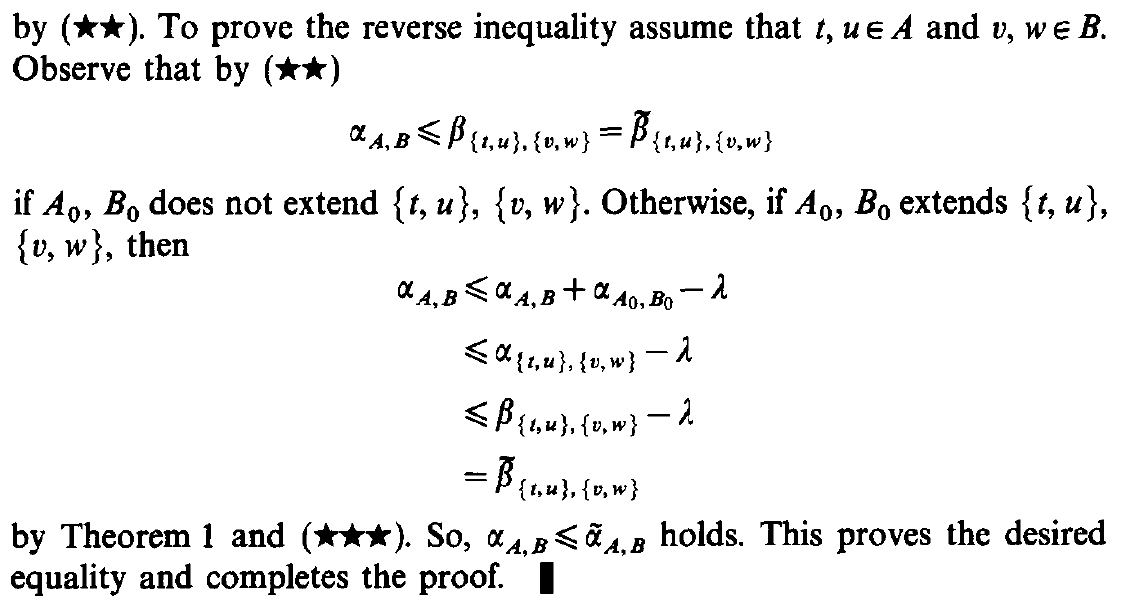
A function is a **dissimilarity function** if

In practice, it is a pseudo-metric without triangle inequality.









**Teo**. 2 Let be a symmetric function.  
Let such that if is a -split  
 and otherwise.  
Then

is a symmetric function such that

In addition, if is a dissimilarity function (or a pseudo-metric),  
then is also a dissimilarity function (or a pseudo-metric).

**Dim**. It suffices to prove the assertions for

where is a -split and .

Then the general case follows by subtracting one split metric at a time (formally induction on the number of non-zero ’s).

We use the following notation:

Clearly is a symmetric function (since and are such).

For every we have

Suppose that is a dissimilarity function.

Then . For and we have

thus , so is a dissimilarity function.

Now suppose that is a pseudo-metric.   
We have to verify the triangle inequality for .

Let . If they all belong to or all to ,  
 then and agrees on and we are done.  
Otherwise, say WLOG and , then we get

where we used the triangle inequality for , and by rearranging

For the remainder of the proof is a symmetric function.

Let be disjoint subsets of .  
We claim that

Suppose .   
Since is a -split

If instead or contains at least three of   
 – say WLOG and – then

So we may assume WLOG that and .  
Since is a -split

from which, similarly as before, we get the inequality

Finally, we claim that for every split

Since , then .

Now let .

If , then .  
So we can suppose and disjoint.

Then choose and such that

We claim that cannot extend .   
In fact, if by absurd it is an extension, then there are two cases.

If , then that implies .  
But is a -split. ⭍

If , then is a -split extending .  
But by **Cor**. 1, such a -split extension is unique,  
 so . ⭍

Since , we get

Now for the same reason, for every and

if does not extend .

Otherwise, if extends , then

where we used **Teo**. 1.

Since

then .

**Cor**. (**split decomposition** / canonical decomposition)

For any symmetric function we can write

where is a split-prime (symmetric) function.  
We call the **split-prime residue** of .

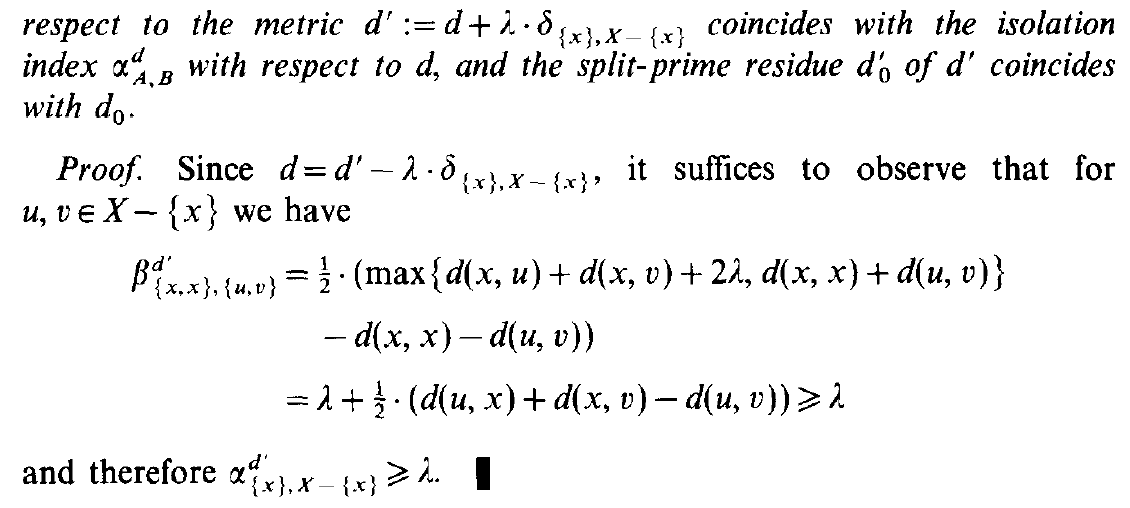
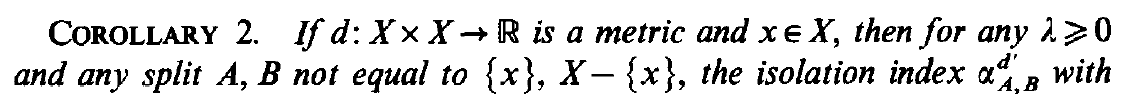
**Dim**. Apply **Teo**. 2 with .

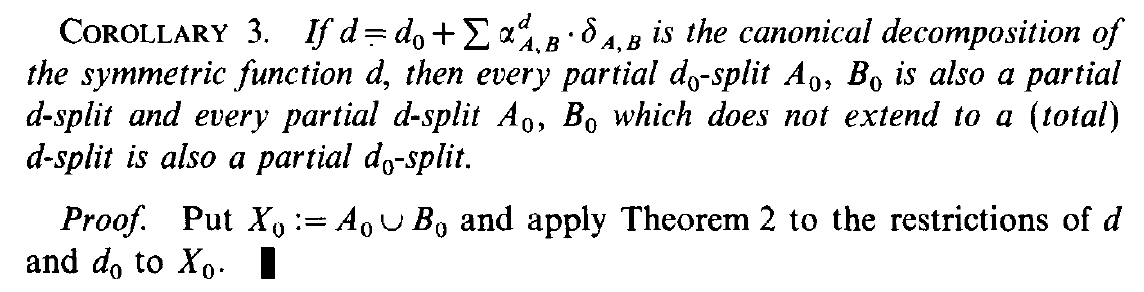
Let .

Then if is a -split, ;  
otherwise .

**Oss**. We can sum over only the -splits, since the others do not contribute (they have zero coefficient).

**Oss**. The residue of a split-prime function coincides with the function itself (there are no splits on which to decompose).





**Cor**. 2 Let be a pseudo-metric and .  
Let and consider

Then we have and

**Dim**. Observe that for every we have

where we used the triangle inequality for .

Therefore and, since ,  
 we get the thesis by applying **Teo**. 2.

Moreover, since , and

where the sums range over the splits different from .

**Prop.** **THIS IS FALSE!!!** The split-prime residue of the restriction coincides with the restriction of the split-prime residue

Moreover, for every partial split of we have

where the sum ranges over all the -splits that extend .

**Dim**. Let . Restricting the canonical decomposition to we get

Instead, the canonical decomposition of the restriction is

But the splits of are the restriction of the splits of to minus the splits such that one of the parts contains .  
For such splits, the split metrics restricted to become the zero function, so we can ignore their splits in the sum.   
Thus the summations in the two decompositions coincide, therefore also the residues.

Servirebbe per tutti gli che estendono .  
Ma questo implicherebbe totally decomposable.

Controesempio: prendi una metrica split-prime e restringerla a un insieme di al più 4 elementi. La sua restrizione è non nulla, ma il residuo della restrizione è nullo.

Consider the case

where is a -split.

We can recycle the proof of **Teo**. 2 with some minor variation.  
First we get that for every disjoint

Suppose that extends . Since

we get .

Now suppose that does not extend .  
Then we prove that cannot extend the quartet that realizes .

In order to apply **Cor**. 1, we need to restrict to

Notice that at least one of these is a total split of ,  
 so it must coincide with .

In order to apply **Teo**. 1, we backtrack in the tree extension to an ancestor at the same level of (observe that this must be different from because is not an extension). Then the sum of their indices is less than the sum of the indices of all the partial splits on the same level, that is less than the index of .

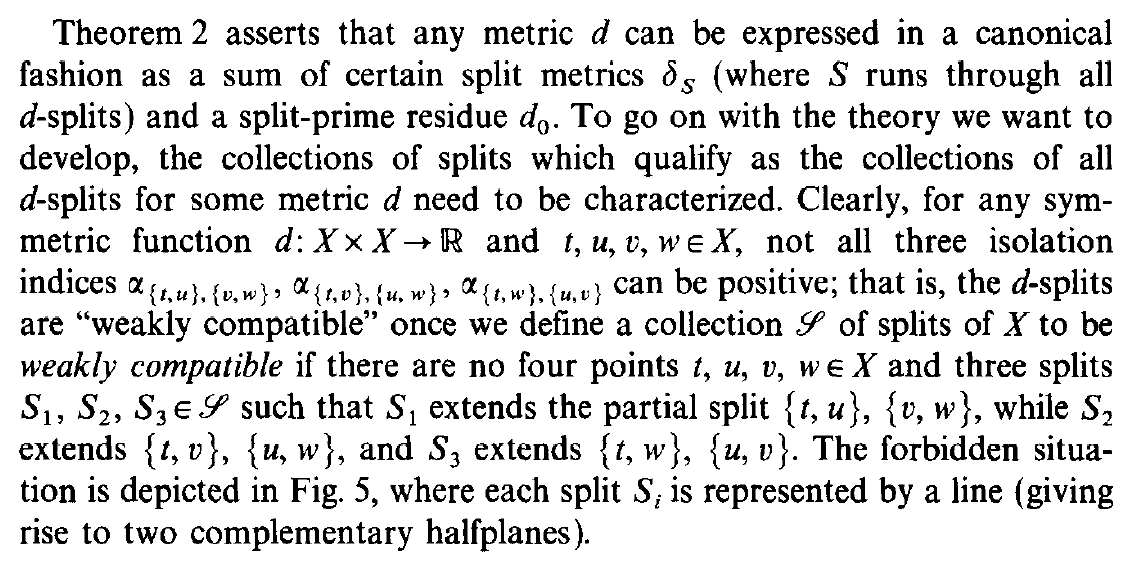
Therefore, for every partial split

We conclude by induction as before.

**Cor**. 3 Let be a symmetric function.

Then every partial -split is also a partial -split;  
and every partial -split, which does not extend  
 to a (total) -split, is also a partial -split.

**Dim**. It is a consequence of the previous **Prop**.



Let be a symmetric function.

**Oss**. Let .  
Then at least one of the following indices must be zero

In fact, suppose

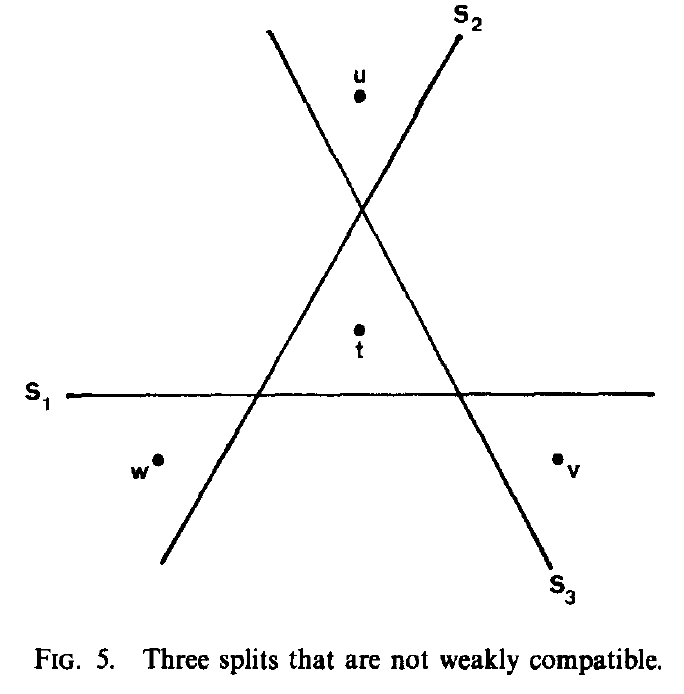
then and so .  
The other cases are analogous.

**Def**. (weakly compatible splits)

Three splits of are **weakly compatible** if  
 there are no four points such that

A set of splits of is **weakly compatible** if  
 its splits are (triplewise) weakly compatible.

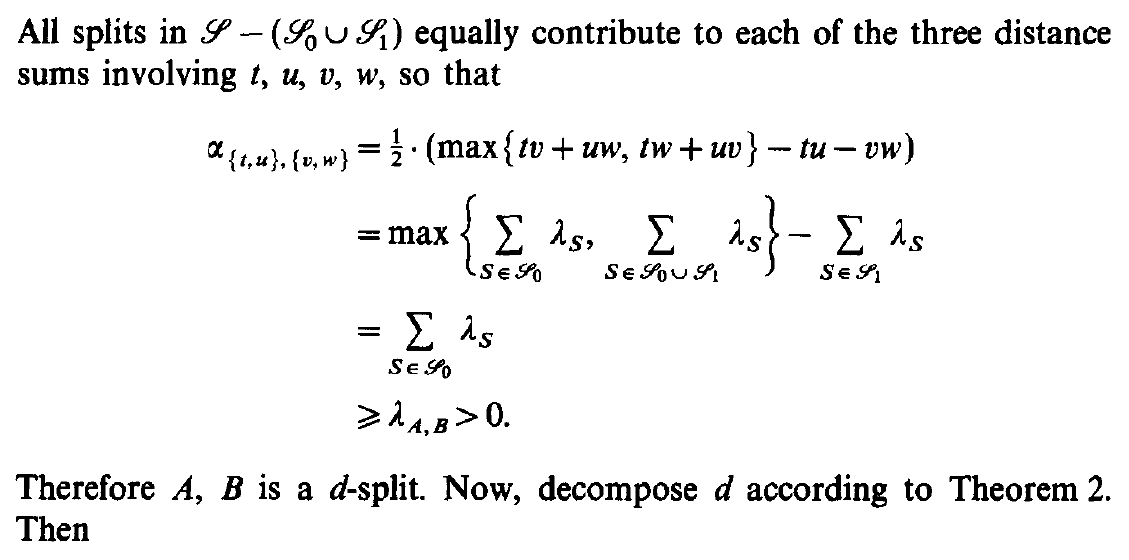
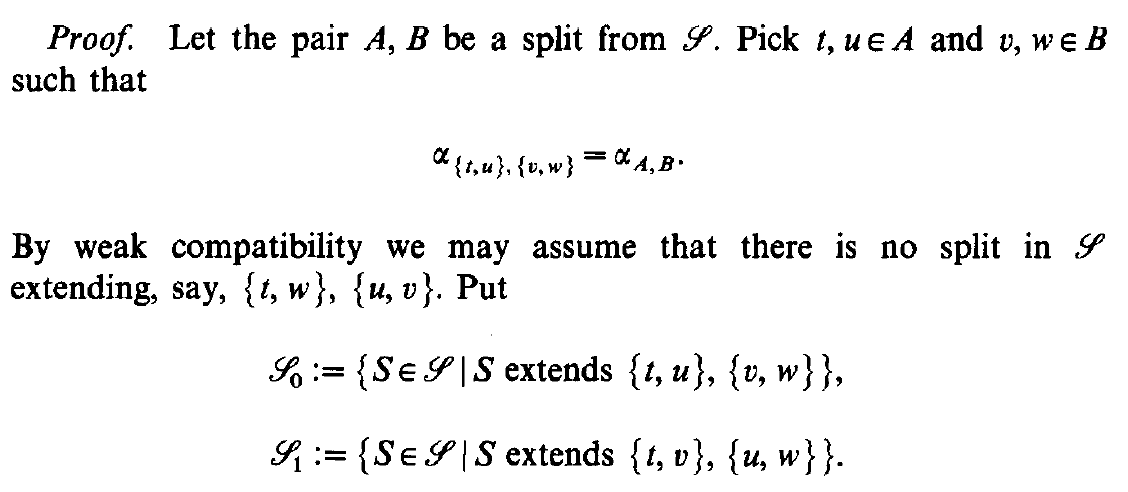
**Oss**. Subsets of weakly compatible sets are weakly compatible.

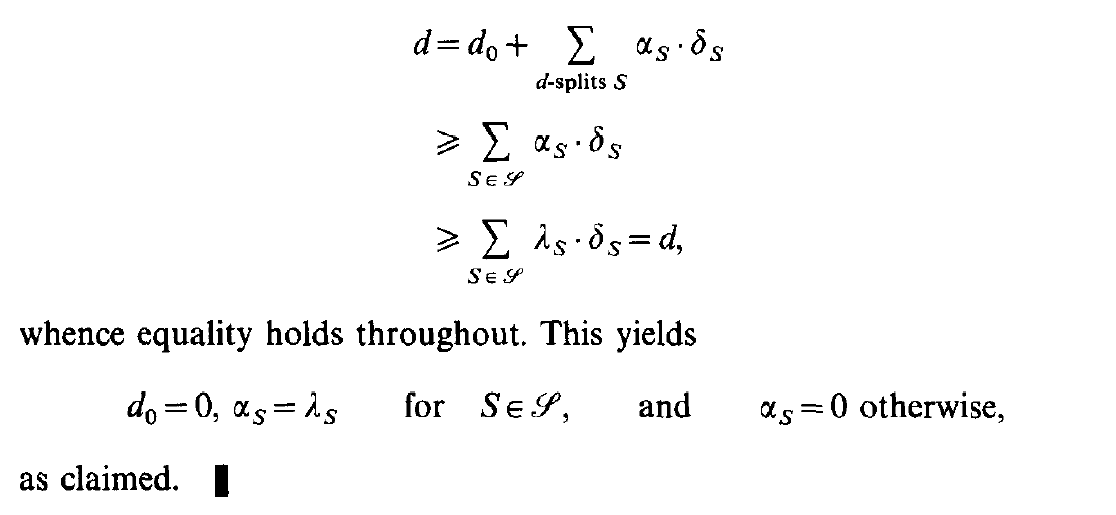
****

**Prop**. The set of all -splits is weakly compatible.

**Dim**. Let . From the previous observation we can suppose WLOG .  
For every split that extends we have

that is is not a -split. In other words, there are no -splits that extend .





**Teo**. 3 Let be a set of weakly compatible splits of .  
For each , let and consider

Then and .

**Dim**. Notice that is a conical combination of split metrics  
 (which are pseudo-metrics), so it is a pseudo-metric.

Let . Pick and such that

Consider the sets

If all three sets are non-empty, then there exist three splits  
 that violates the weak compatibility assumption.   
So at least one of them is empty, say WLOG .

All splits in equally contribute to each of the three distances ;  
so we can ignore them in the calculation of the isolation index:

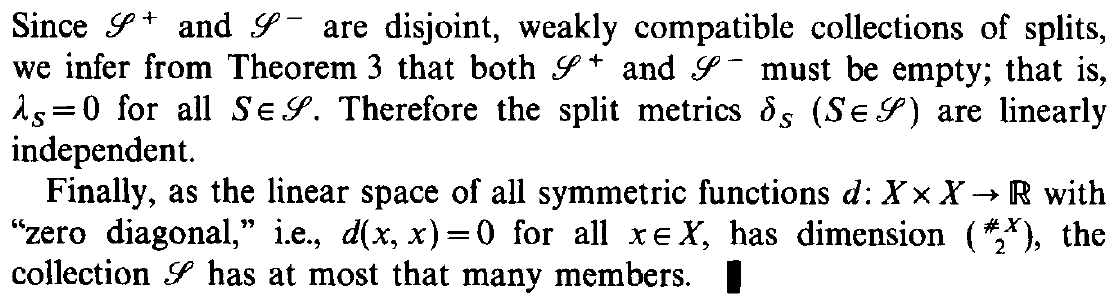
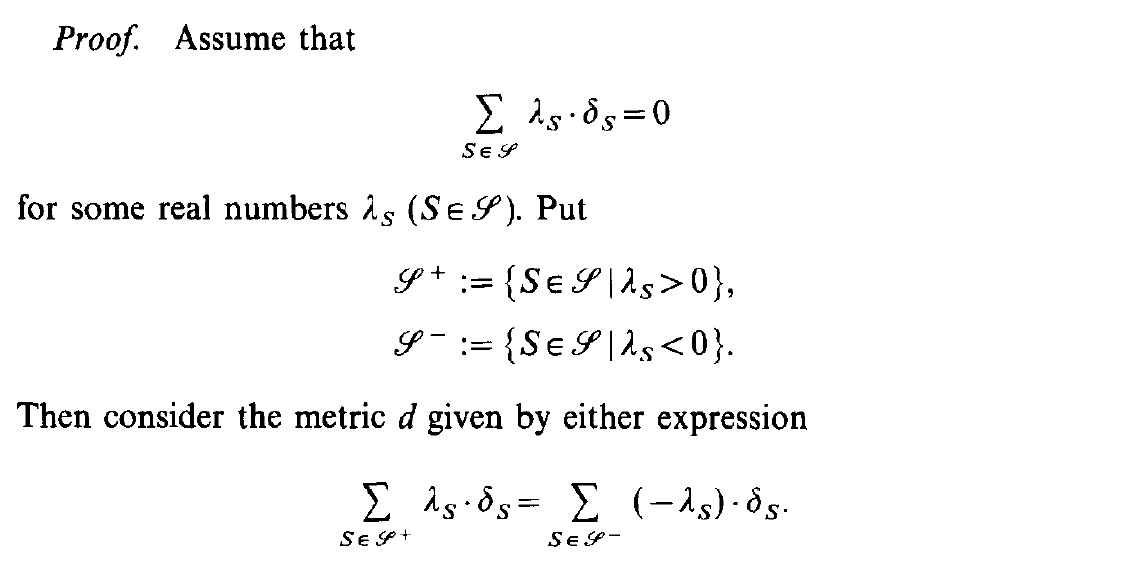
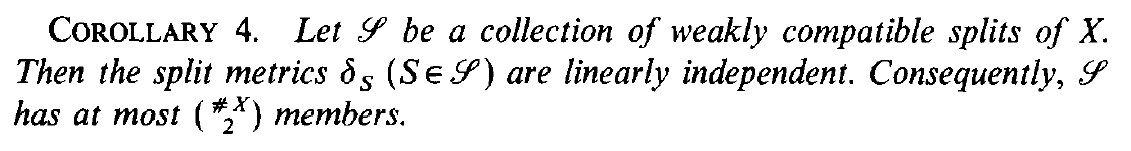
So for every we have   
 and is a -split, therefore .

Decomposing as in **Teo**. 2

where we used that , since it is a pseudo-metric.  
So the inequalities hold as equal. Thus

that is if , then is not a -split.

We conclude that .



**Cor**. 4 Let be a set of weakly compatible splits of .

Then the split metrics are linearly independent.  
Also , where .

**Dim**. In order to prove the linear independence, suppose

for some for each .

Let

We can decompose the previous expression in

Consider the pseudo-metric

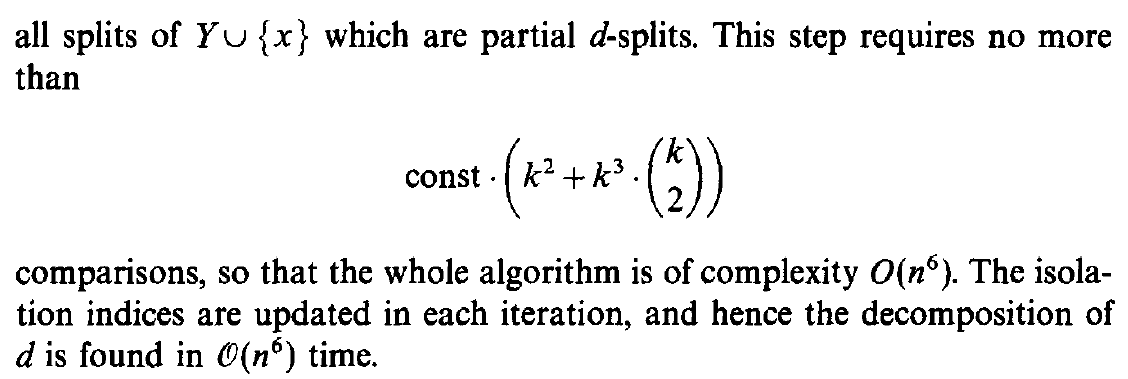
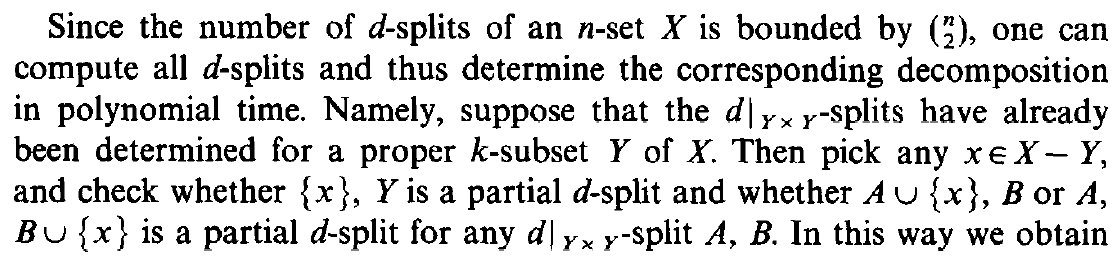
Observe that both and are weakly compatible.  
Applying **Teo**. 3 to the first expression of we get

and doing the same with the second expression we get

So .  
But and are disjoint due to how they are defined.  
So they are both empty: .

We conclude that .  
Therefore the split metrics are linearly independent.

Since and then



**Oss**. Since is weakly compatible, from **Cor**. 4 we have that the number of -splits is at most .

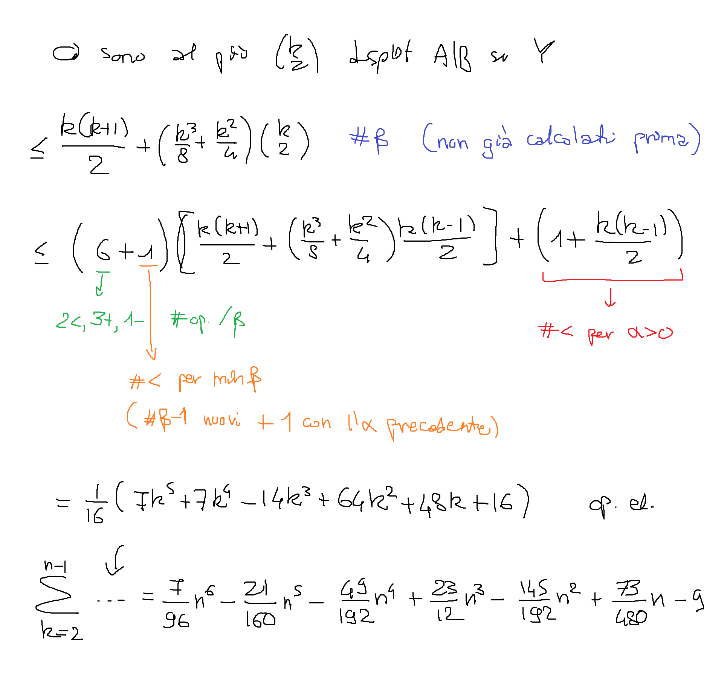
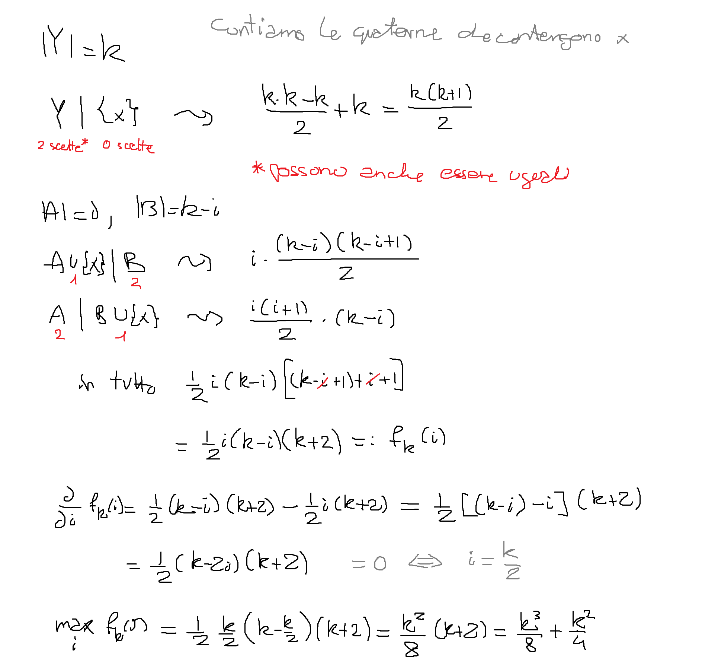
A brute force approach to find all the -splits of would be to compute the isolation indices of all the splits of and discard those that are zero.  
But, since , this is an exponential algorithm.

We can instead use a more “inductive” approach:   
suppose that the -splits of a proper subset of size have already been determined. Then pick any and check

* if is a partial -split
* if and are partial -split  
   for any -split of

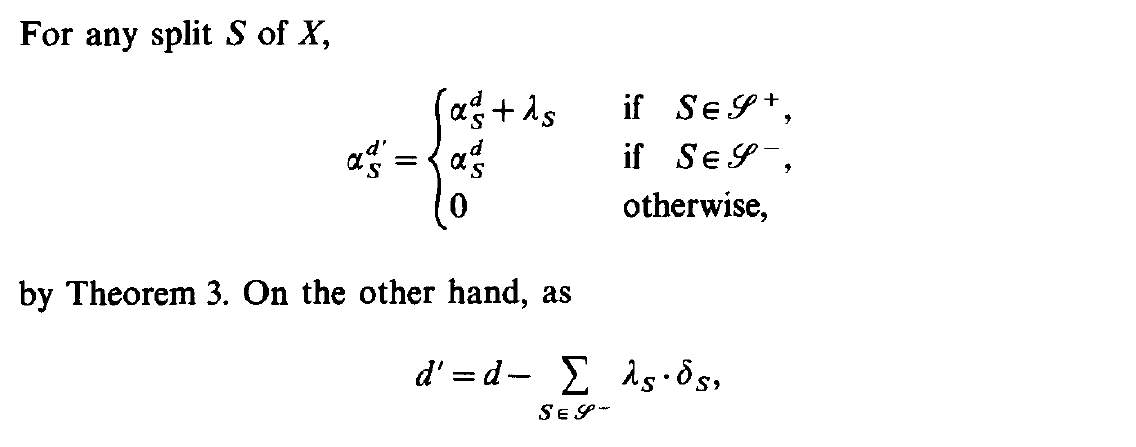
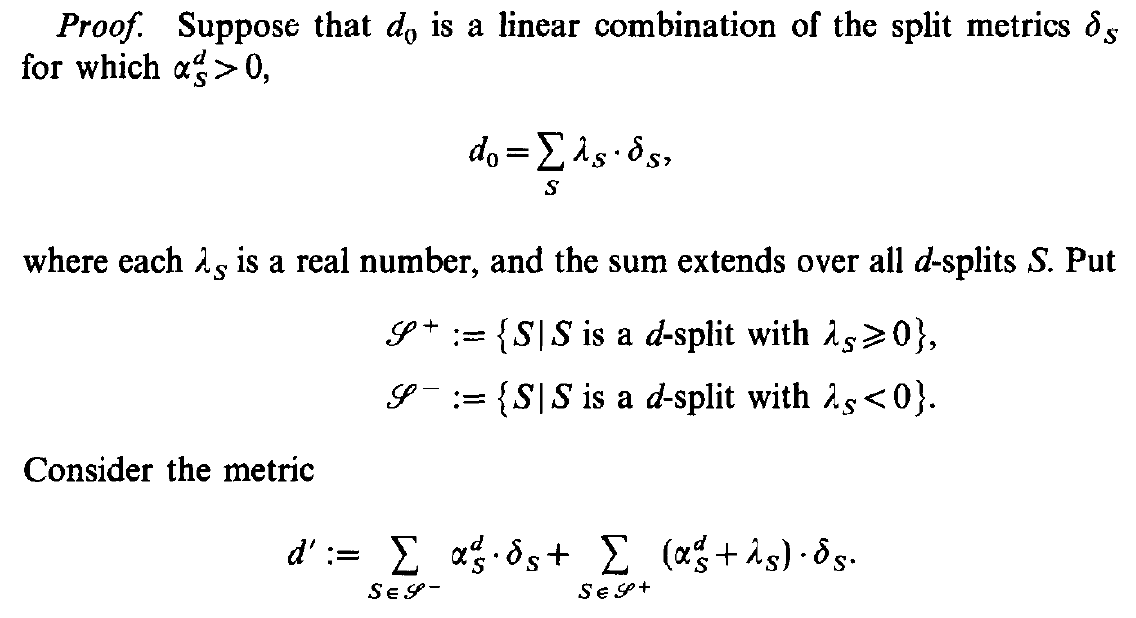
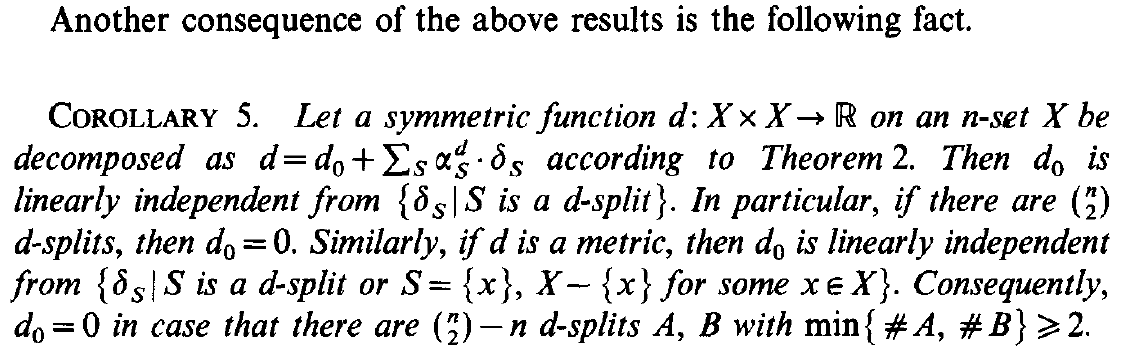
In this way we obtain all the -splits of .

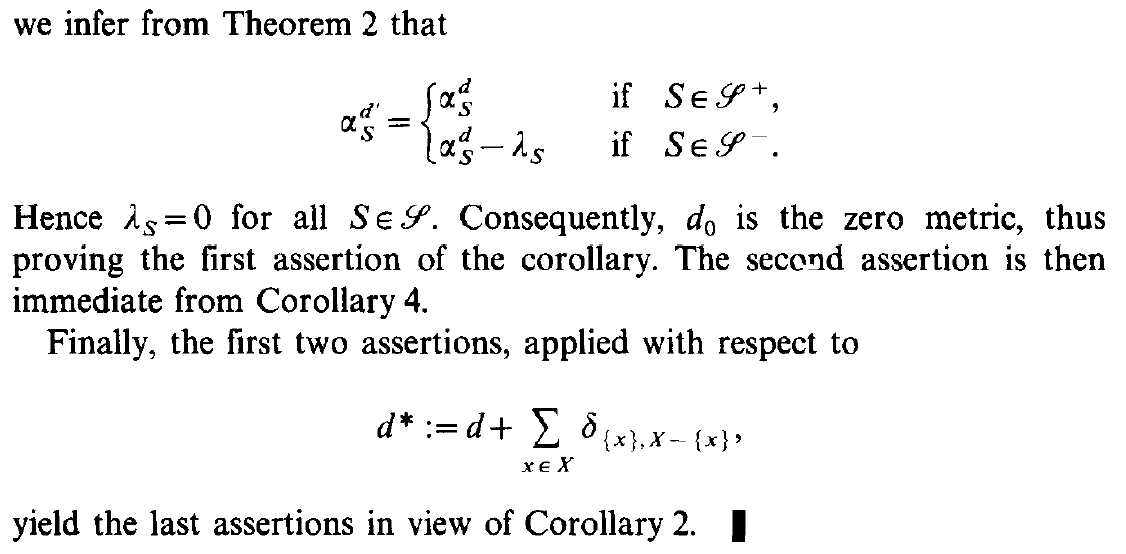
Crucially, we just have to check at most splits at each step, thanks to the previous observation. So we can compute the -splits of (and their decomposition) in polynomial time.



The reason why the algorithm works is that once we find splits that are not partial -splits (that is they have zero isolation index), we know that also their extensions won’t be -splits – since the isolation index can only get lower by extending.

A similar improvement in the actual implementation can be made: once we find that a quartet has zero index, then also the index is zero;   
so we can stop checking the other quartets.  
Also when computing the index we can exploit the symmetry of to skip about of quartets.





**Cor**. 5 Let be a symmetric function.

Then the residue is linearly independent from .  
In particular, if there are -splits, then .

If is a pseudo-metric,  
 then is linearly independent from .  
If there are non trivial -splits, then .

**Dim**. Suppose that

so that

Let

Observe that .

Consider the pseudo-metric

Applying **Teo**. 3 we get

We can write

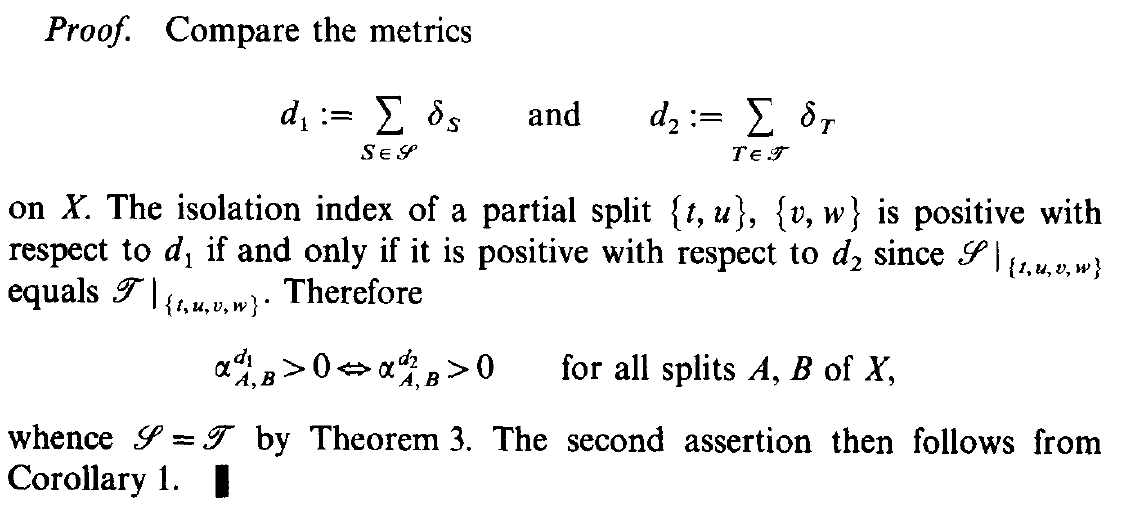
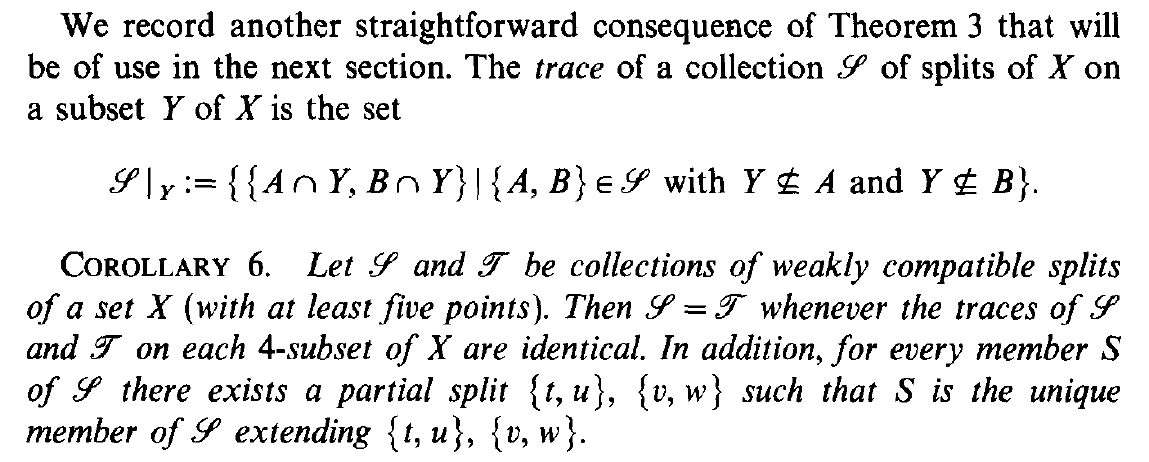
and by applying **Teo**. 2 we get

Thus for every split , proving the linear independence. The second assertion follows from **Cor**. 4.

Consider

By **Cor**. 2, . Also notice that

We get the thesis by applying the first assertion on .



**Def**. Let be a collection of splits of and a subset.  
Then the **trace** of on is the set

In practice, it is the restriction of the splits to such that they are still splits: in particular the members cannot be empty and, since the set of all splits is a partition of , it is equivalent to ask that is not contained in a single member.

**Cor**. 6 Let and be collections of weakly compatible splits of .

Then if and only if   
 their traces are identical on every 4-subset of .

**Dim**. The implication is obvious.

Consider the pseudo-metrics  
 (since they are conical combinations of split metrics)

Observe that, given a 4-subset of ,

because they depend only by (the restrictions of) the splits on , and have the same trace on 4-subsets.

If we consider a split on , for example ,   
and the fact that are pseudo-metrics, we have

In particular, the isolation index on quartets is positive with respect to if and only if it is positive with respect to .

This is true also for generic splits because  
 the isolation index on a generic split is the minimum of  
 the isolation indices on appropriate quartets.  
In particular, the -splits coincide with the -splits.

By **Teo**. 3 we have

**Prop**. Let be a collection of weakly compatible splits of .

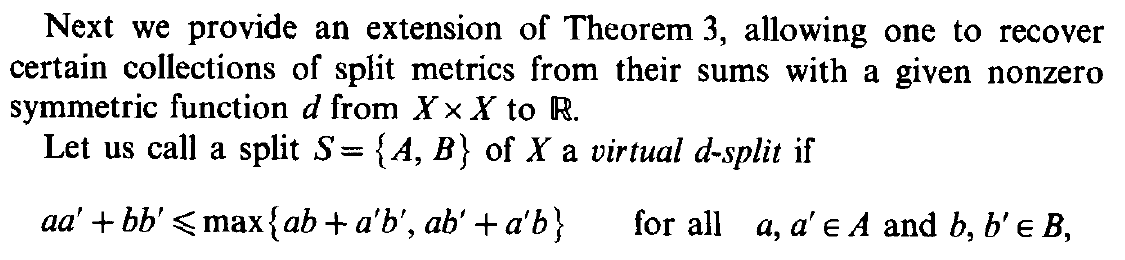
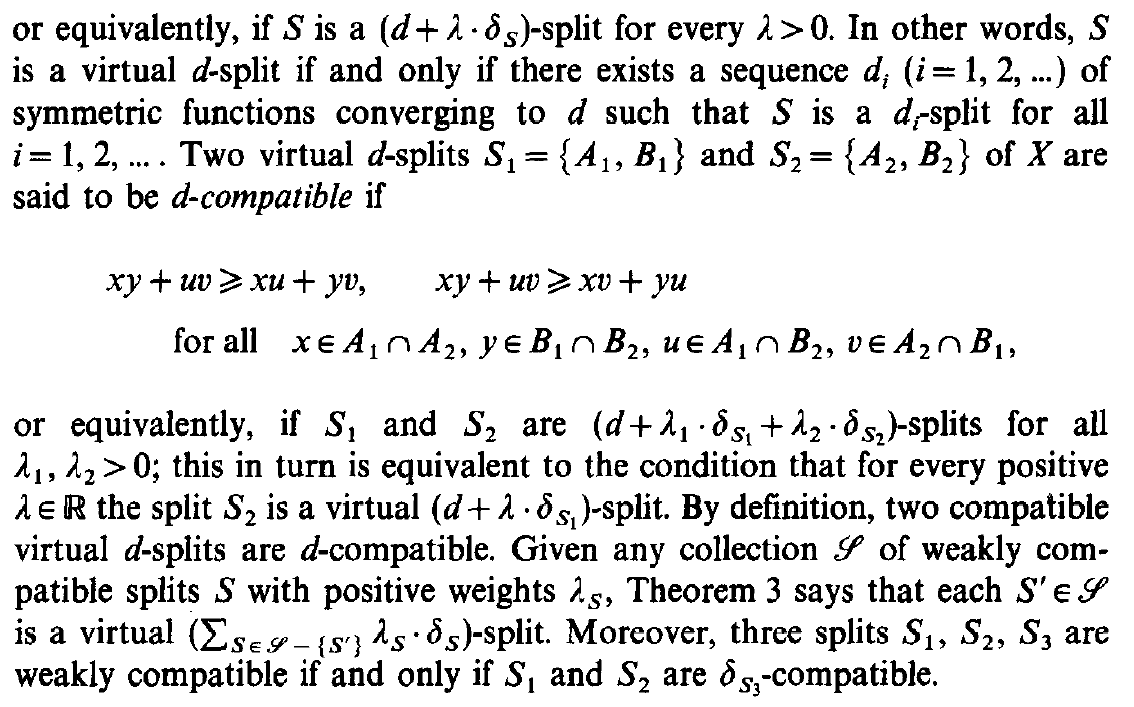
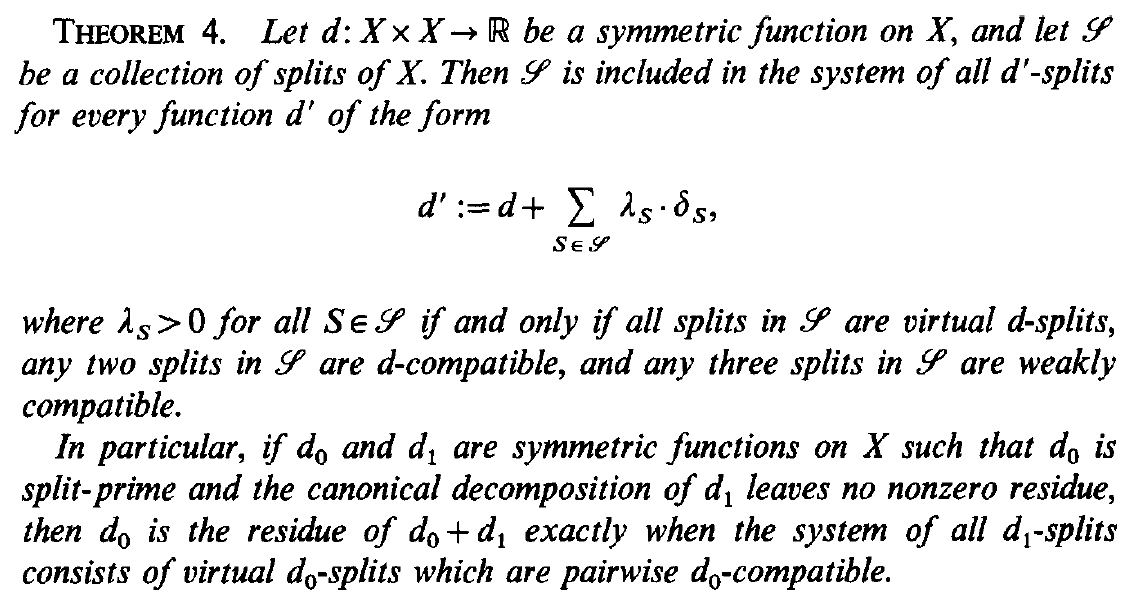
Then for every there exists a partial split such that is its unique split extension in .

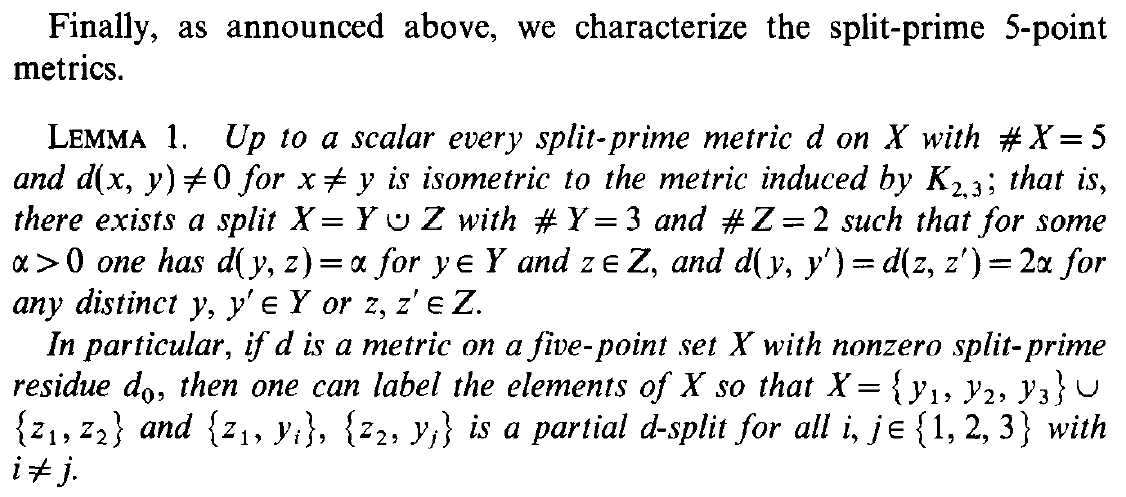
**Dim**. Let such that

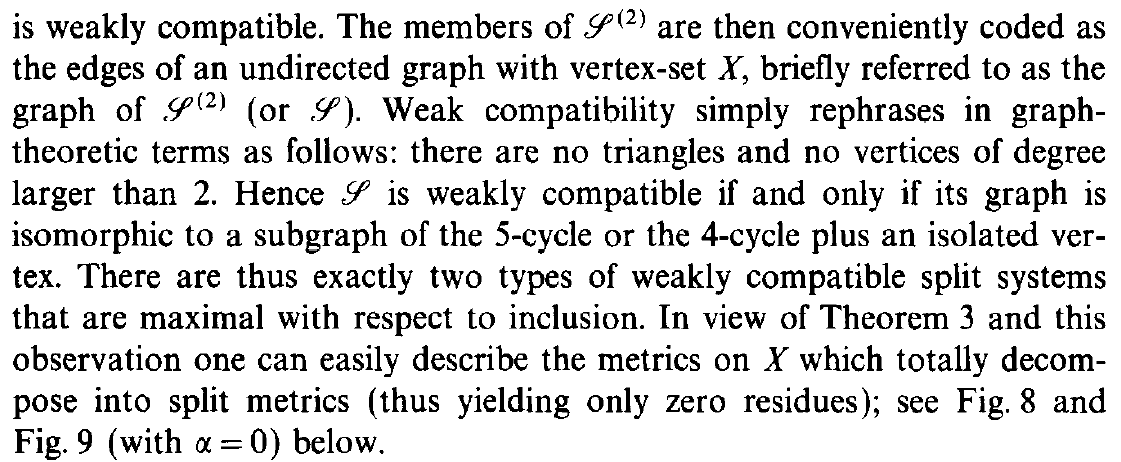
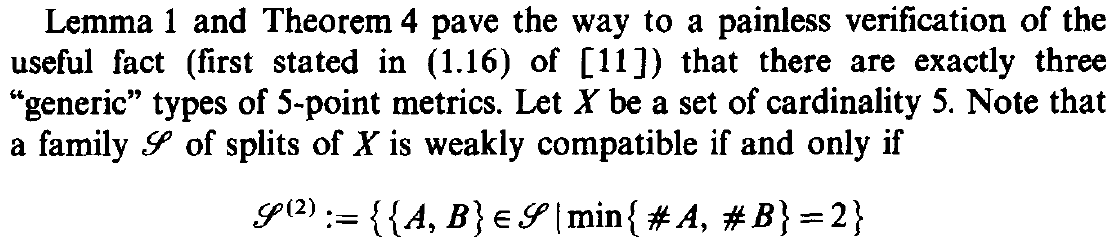
with respect to the distance .

Since by **Teo**. 3 it holds , then is a -split.  
Applying **Cor**. 1 we conclude that is the unique -split (that is equivalent to saying element of ) extending .

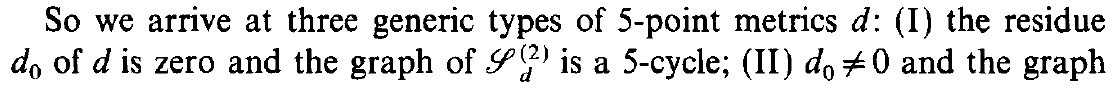
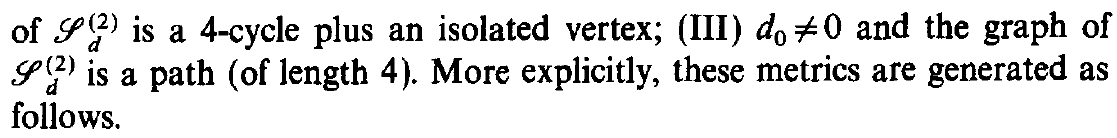
Solo enunciati.

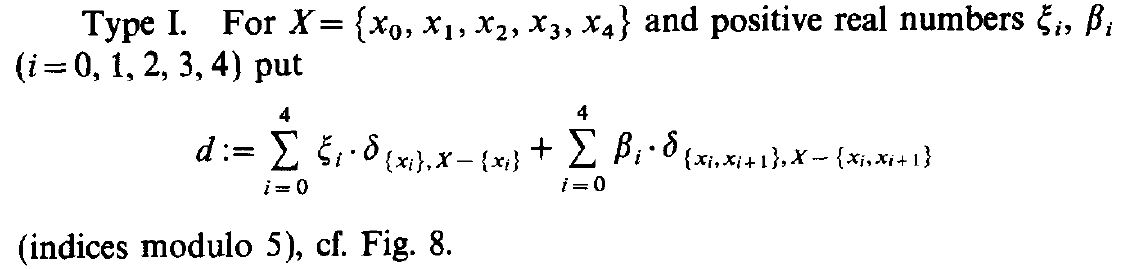
  

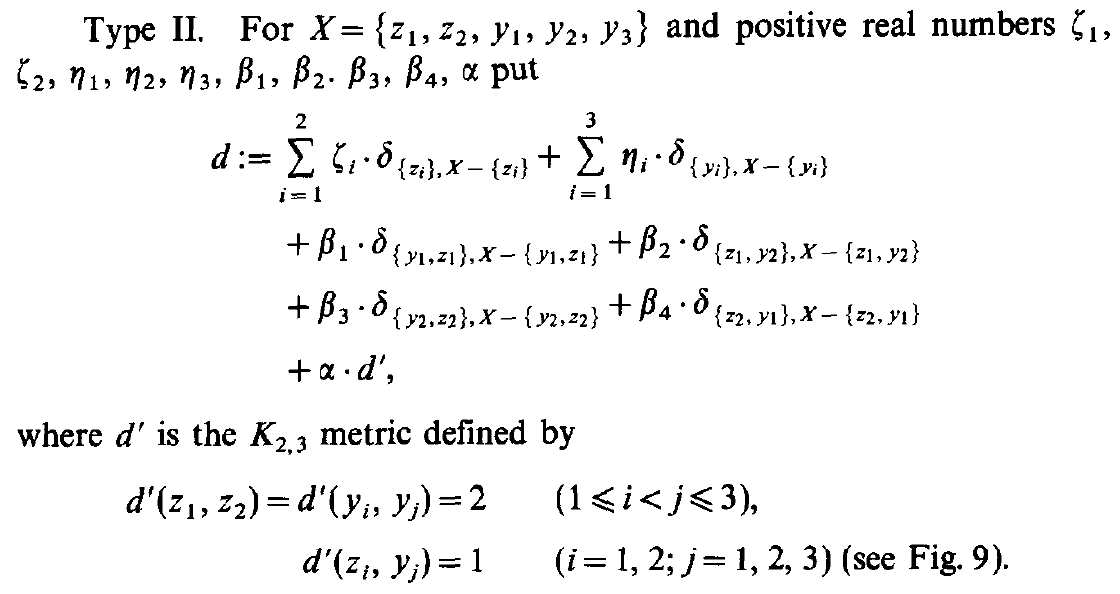


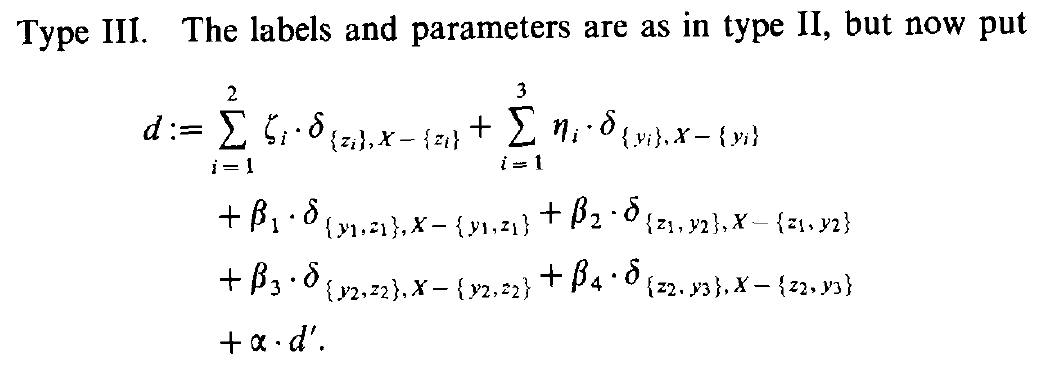


[...]

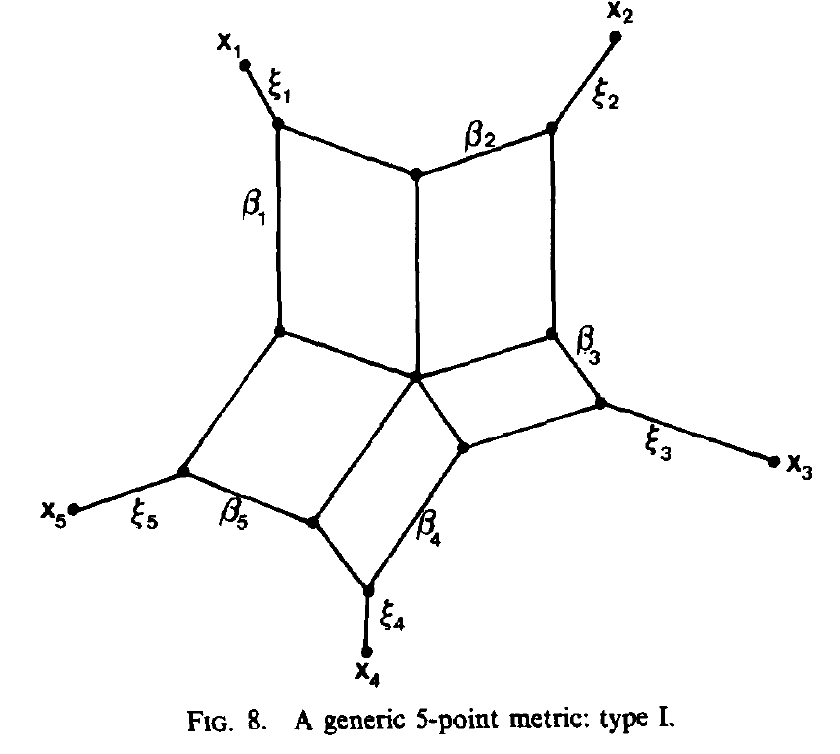
 

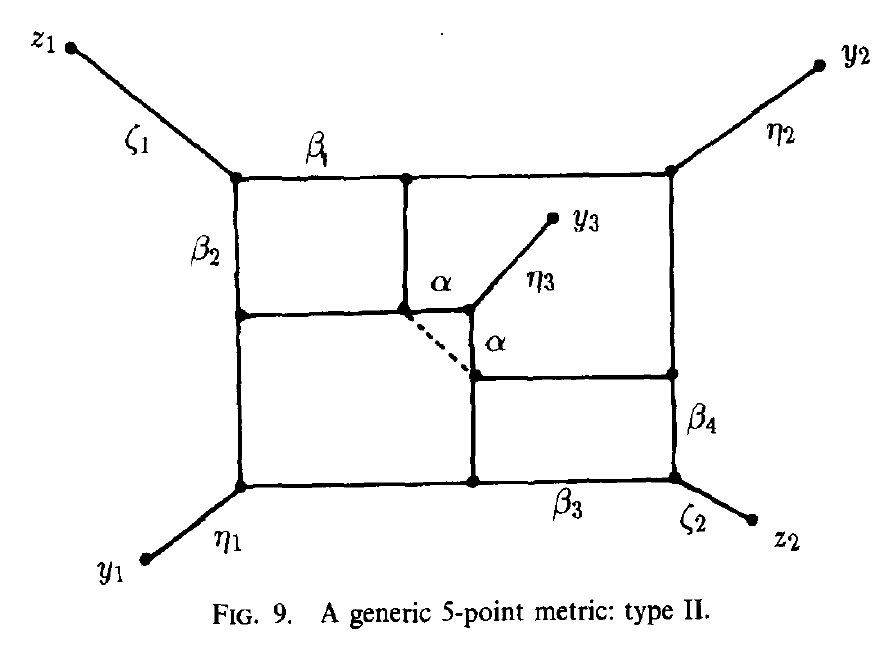


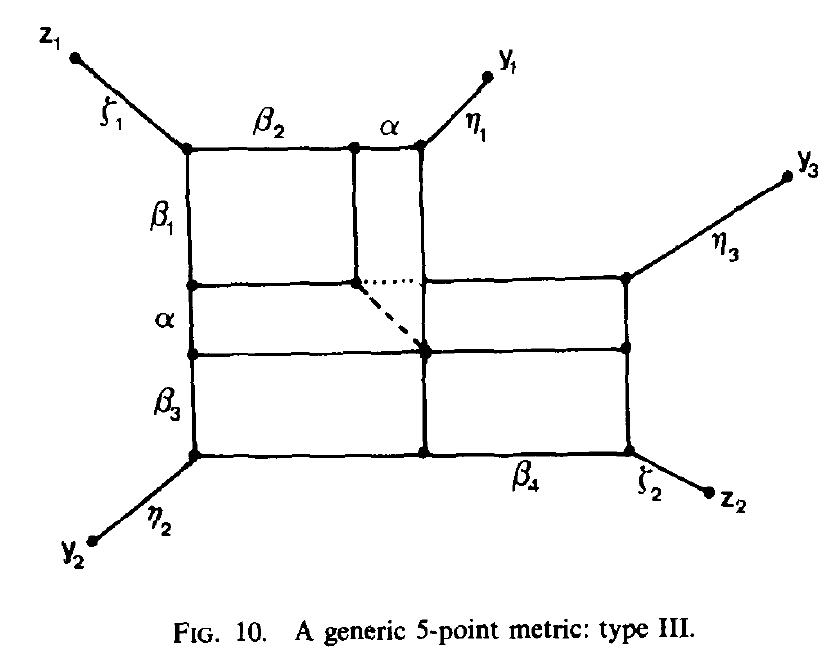




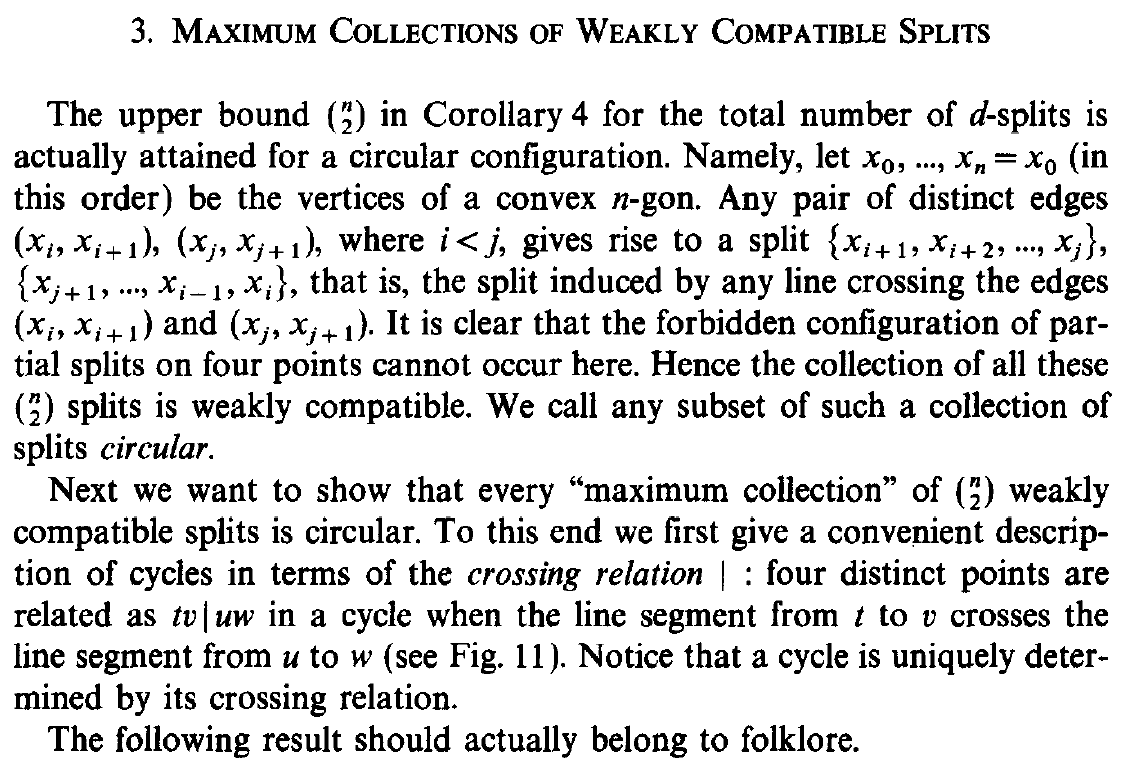


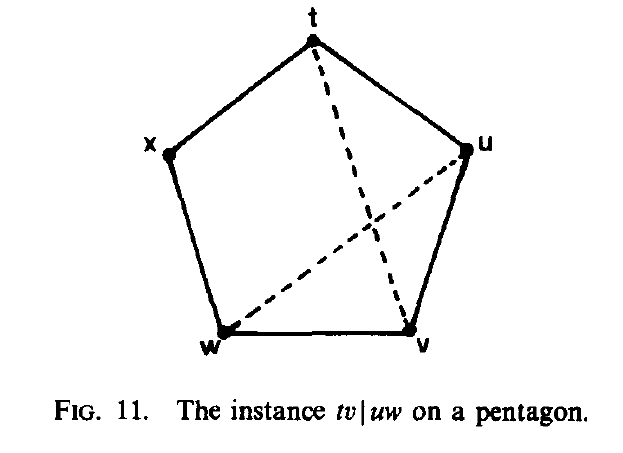




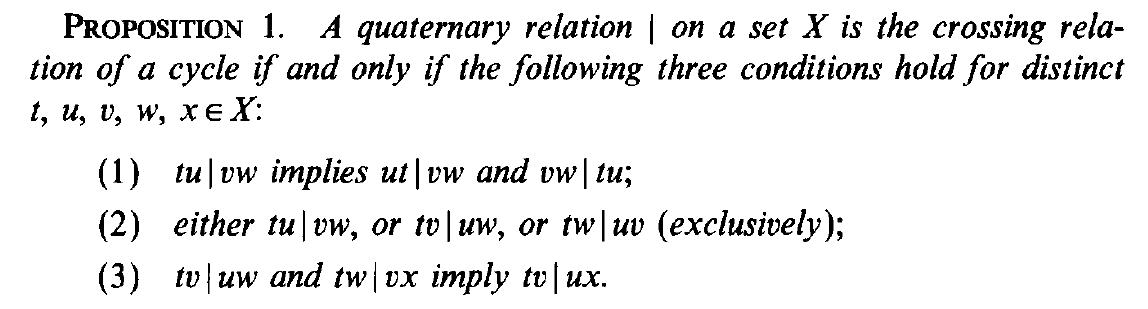


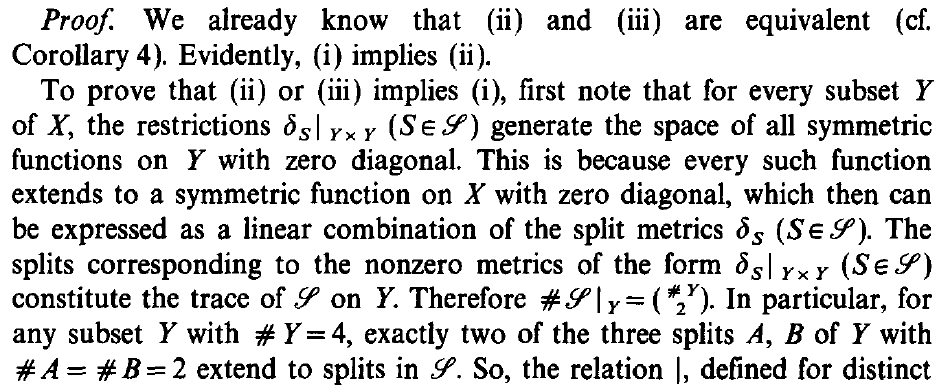
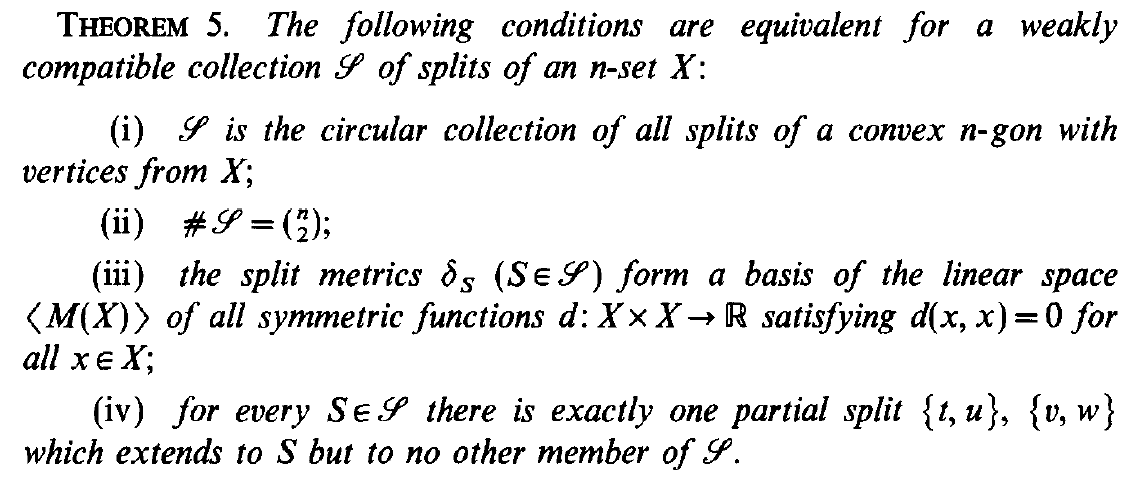
Chapter 3: Maximum Collections of Weakly Compatible Splits

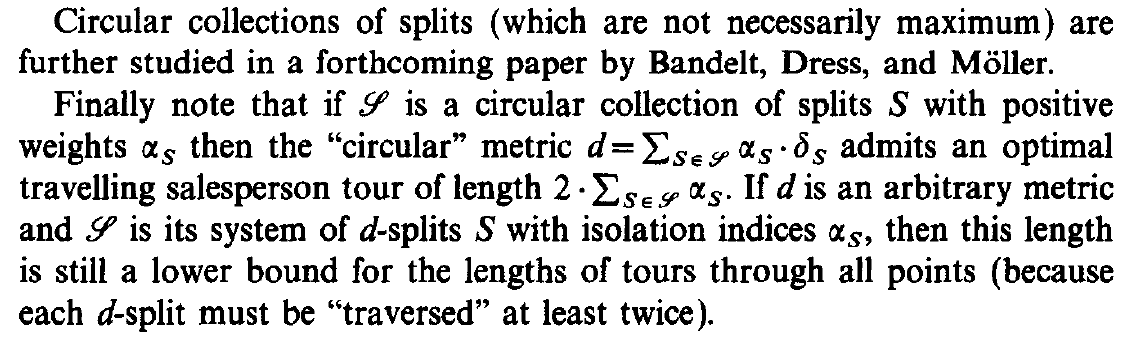




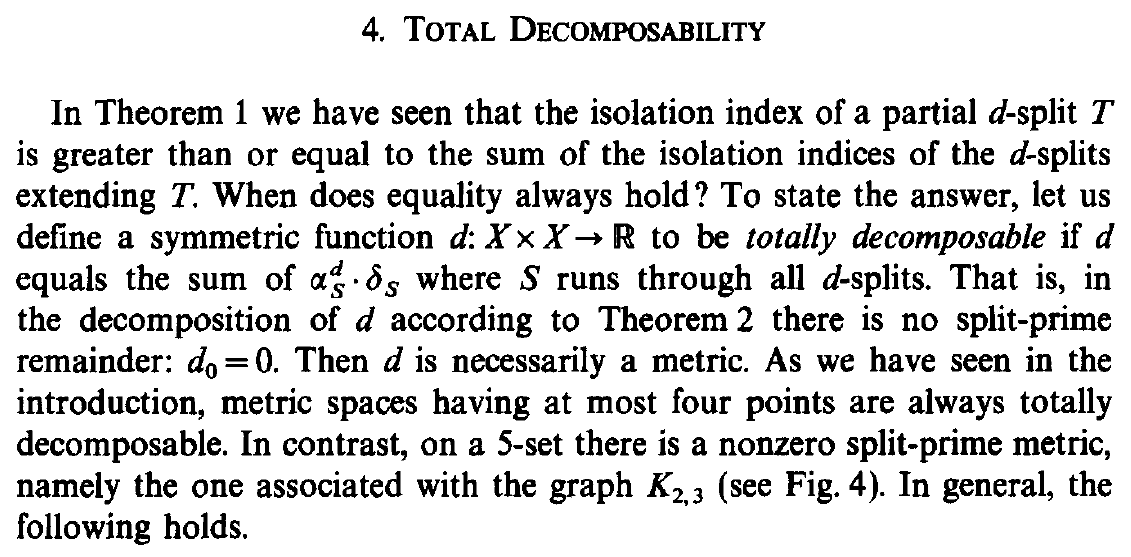
I circular splits sono weakly compatible perché 4 punti disposti in cerchio non sono linearmente separabili.







Chapter 4: Total Decomposability



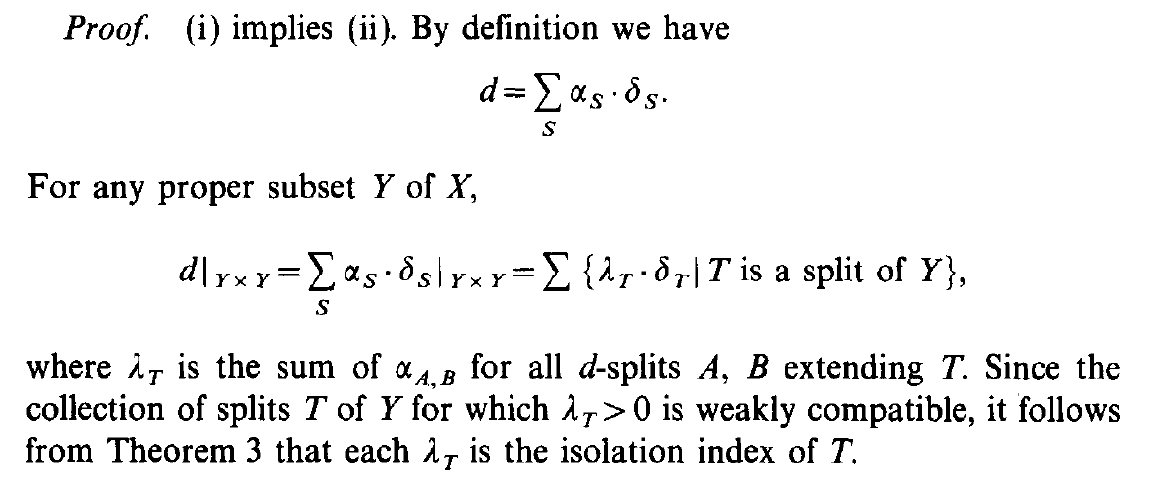
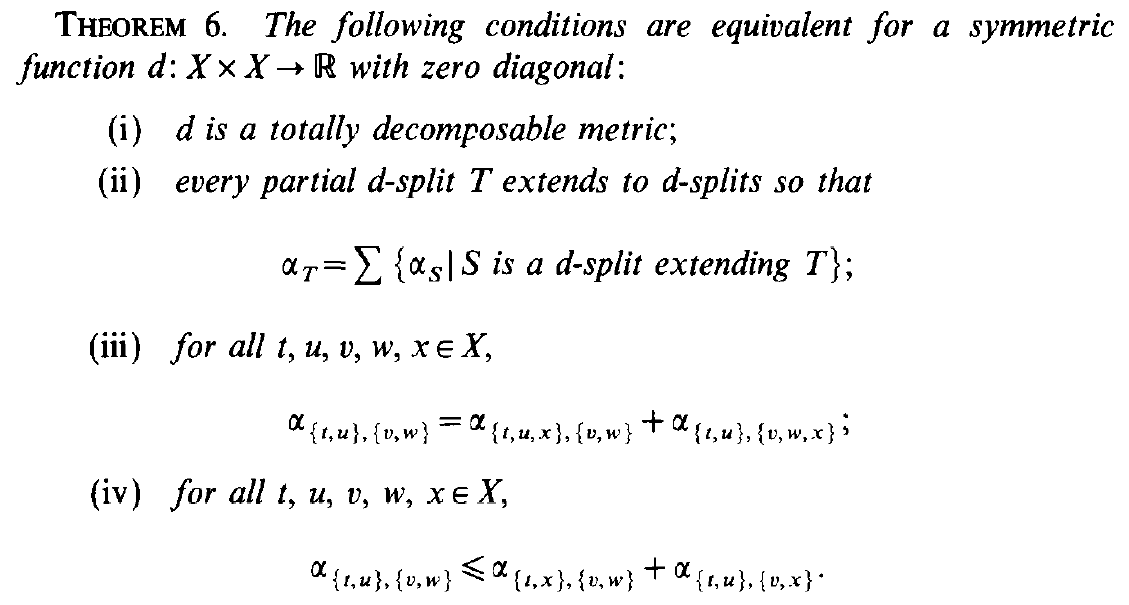
**Def**. A symmetric function is **totally decomposable** if its split-prime residue is zero ;   
or equivalently, if it can be written as

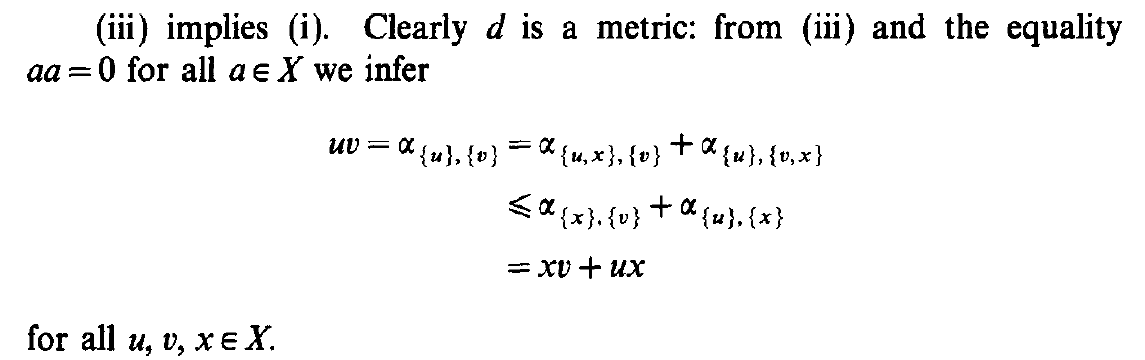
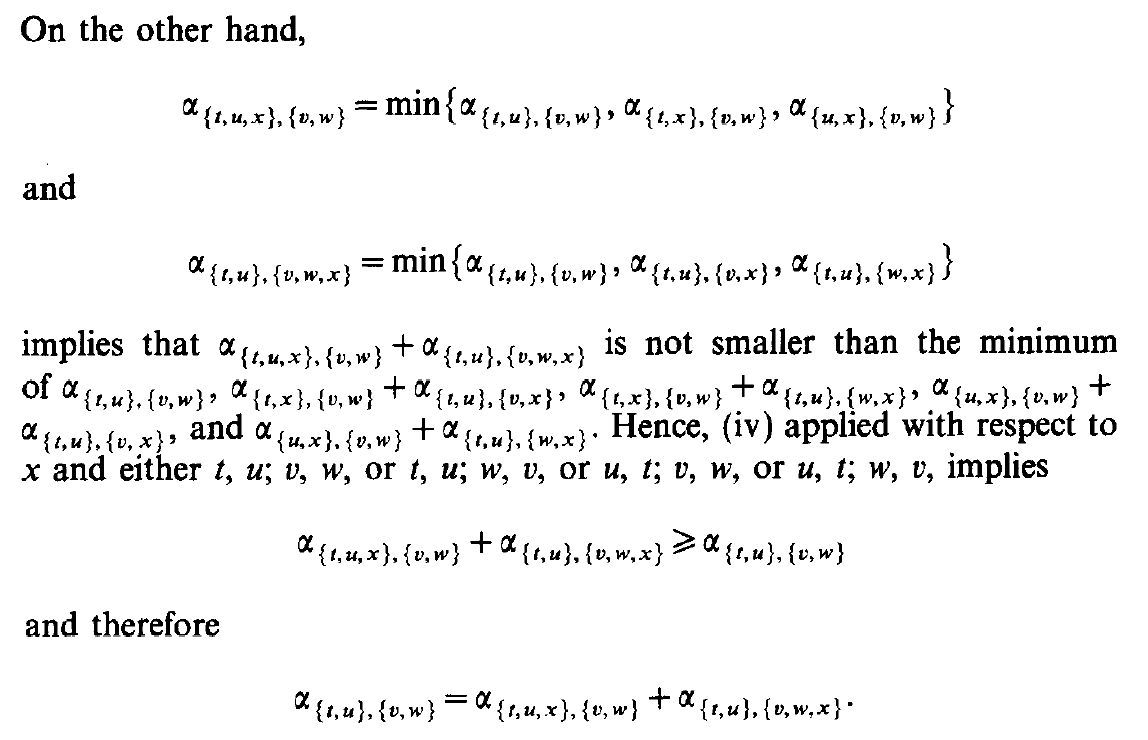
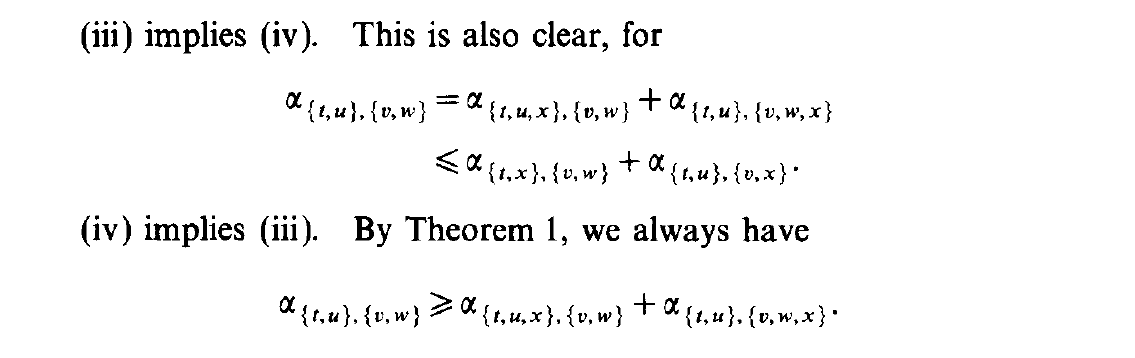
**Oss**. A totally decomposable function is a pseudo-metric.

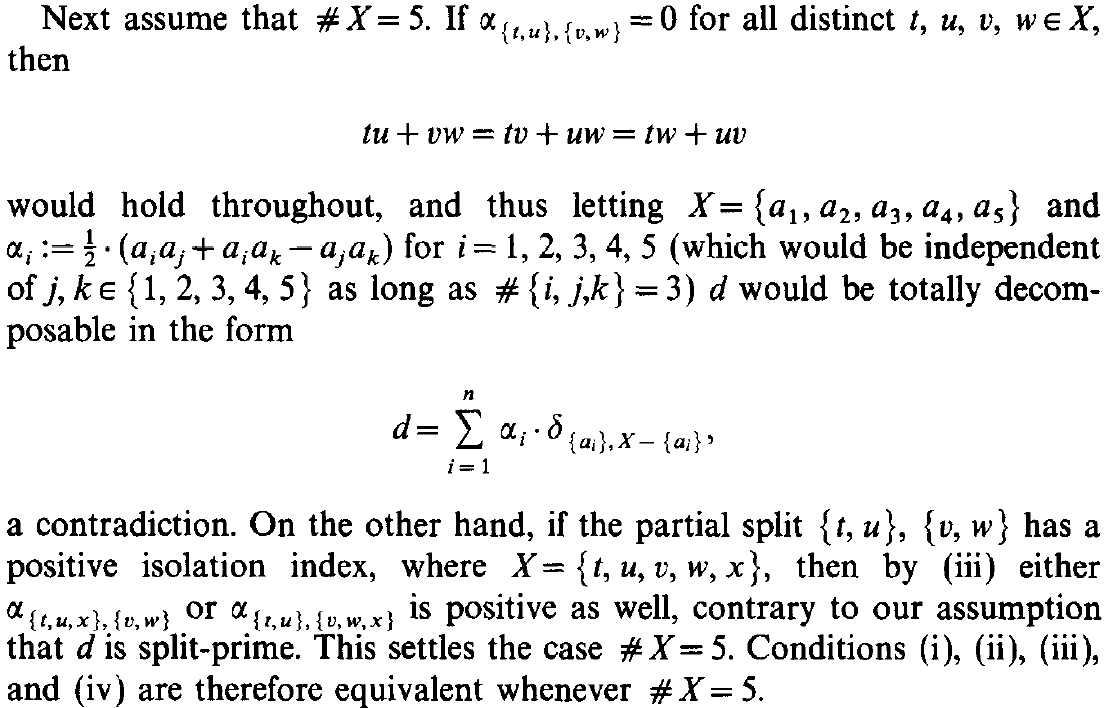
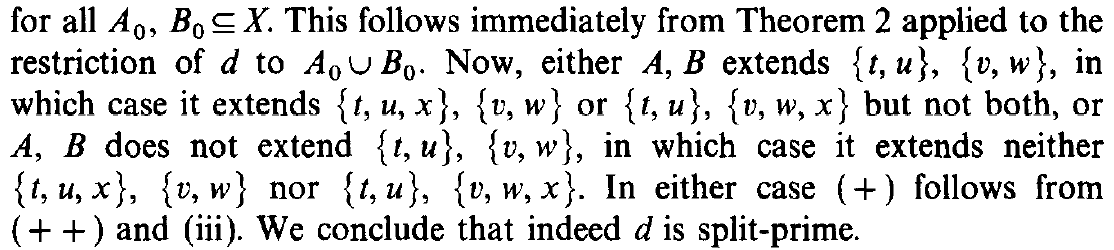
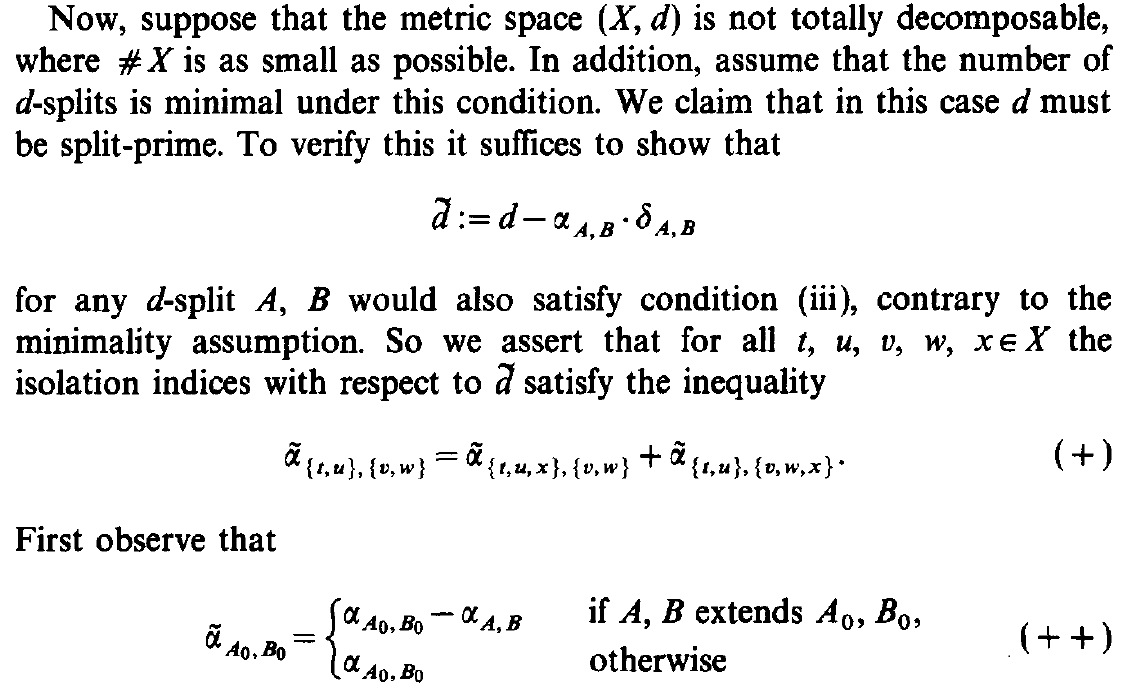
In fact, it is a conical combination of split metrics  
 (that are pseudo-metrics).

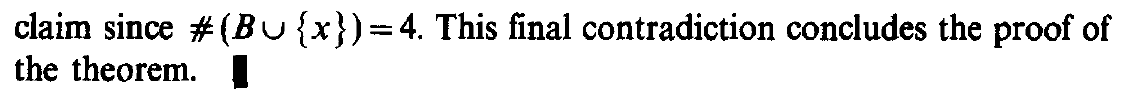
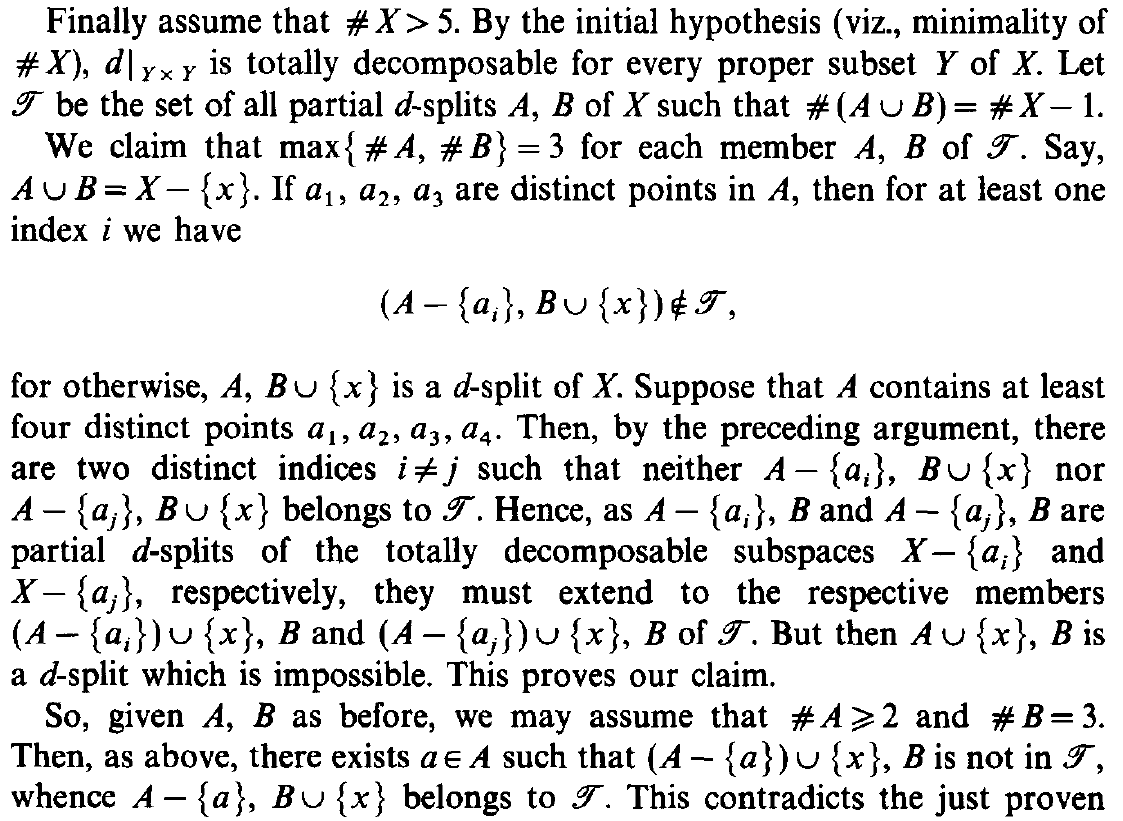
**Fact** If , then every pseudo-metric is totally decomposable.

If , then there is a non-zero split-prime pseudo-metric, namely the one associated with the graph .









**Teo**. 6 Let be a symmetric function with zero diagonal. Then the following conditions are equivalent:

1. is totally decomposable
2. for every partial split
3. for all
4. for all

**Dim**. By definition of total decomposability, we have

For any proper subset

where .

From **Teo**. 1, we have for every split of . Thus

so it is weakly compatible. Applying **Teo**. 3 to this set we get

Let . Then

Let . Then

Let . By **Teo**. 1 we have

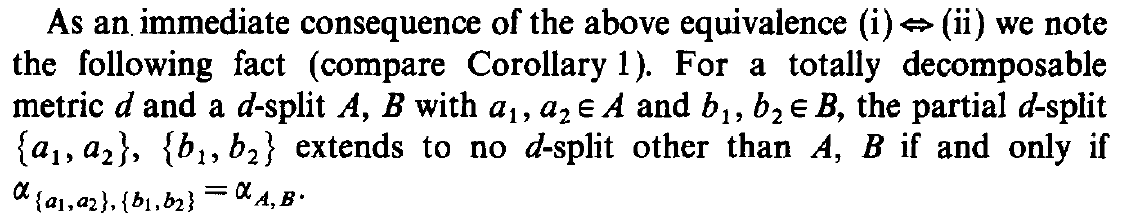
On the other hand

Applying condition with respect to and either

we get

Therefore

[DA FARE]



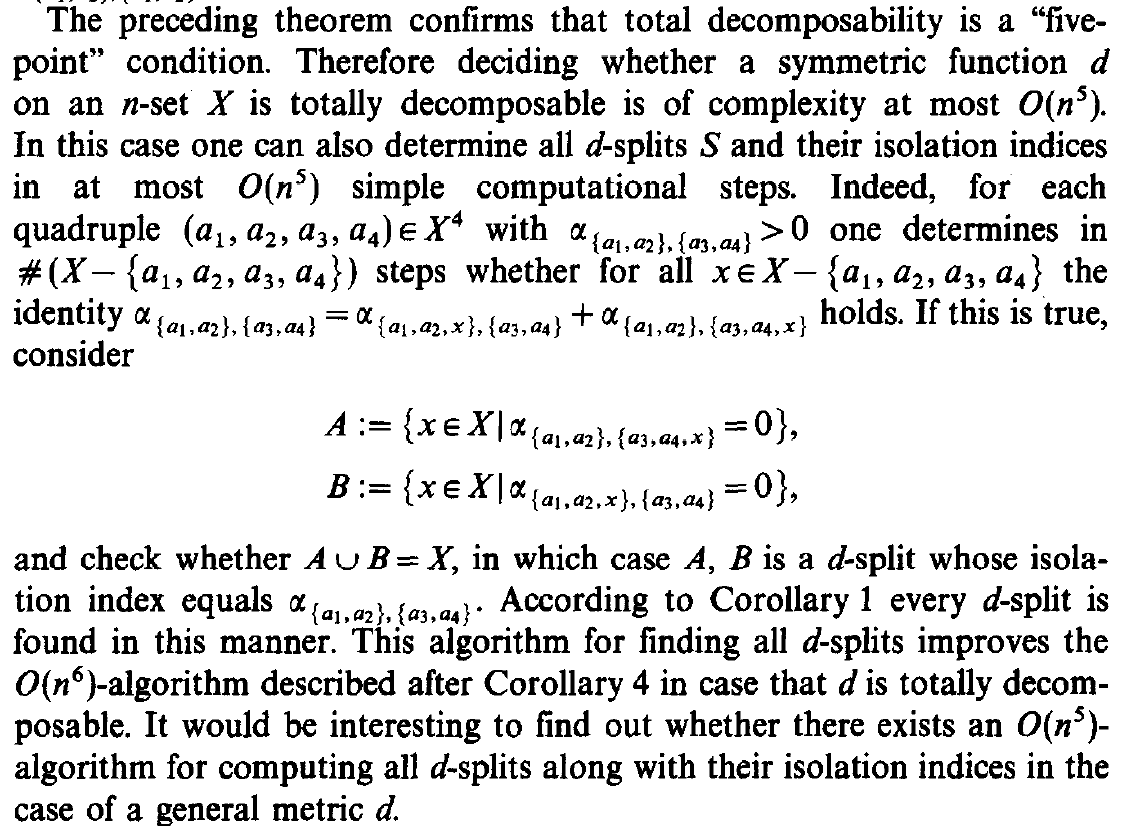
**Cor**. Let a totally decomposable pseudo-metric,  
 a -split and .

Then is the only -split extension of   
if and only if .

**Dim**. By **Teo**. 6

By **Teo**. 6

and since is a term of the sum, it must be the only one.



Consider such that and the sets

Suppose that the following identity holds for all

Then we have and .

Since and, by extending the (partial) splits, the isolation index cannot increase, then all extensions of with at least one element of in the second part have isolation index equal to 0.  
Idem for extensions of with at least one element of in the first part.

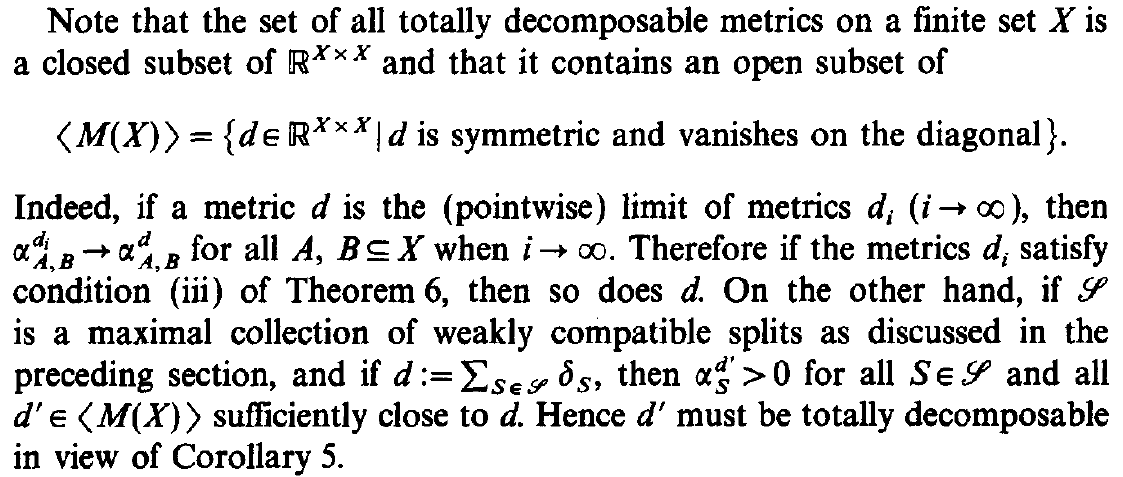
So between the extensions of , the only possibly non zero isolation index is that of the split (that is all elements of in the first part and all elements of in the second one).

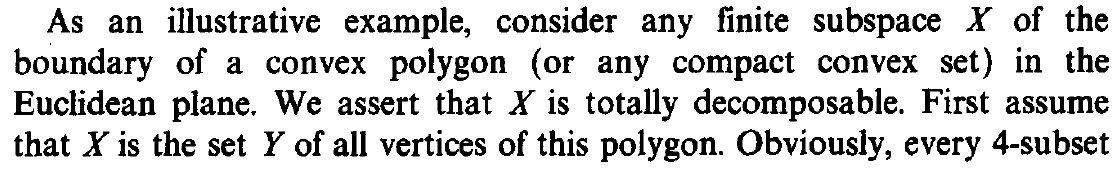
If is totally decomposable, then the sum of the isolation indexes of the extensions equals the isolation index of the base partial split, and so

This gives an algorithm to check whether a symmetric function is totally decomposable and to compute -splits and their isolation indexes:

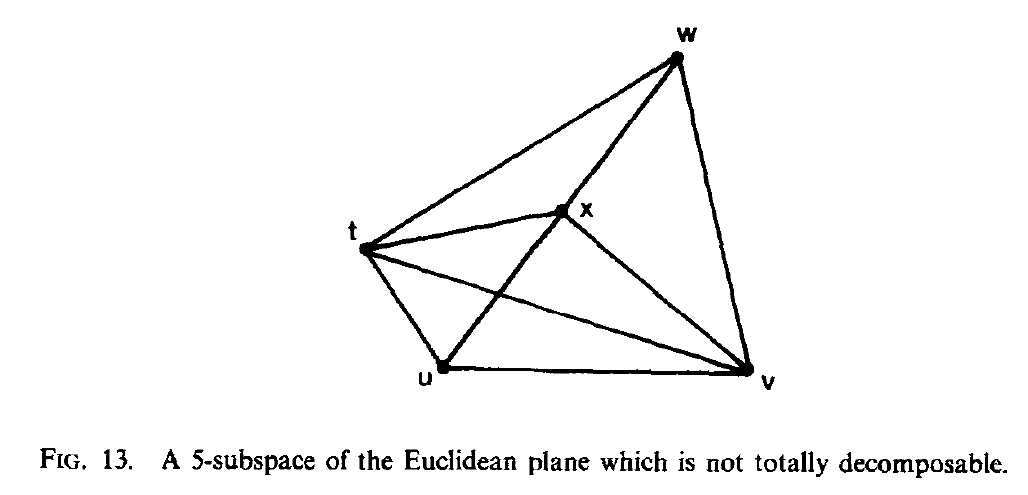
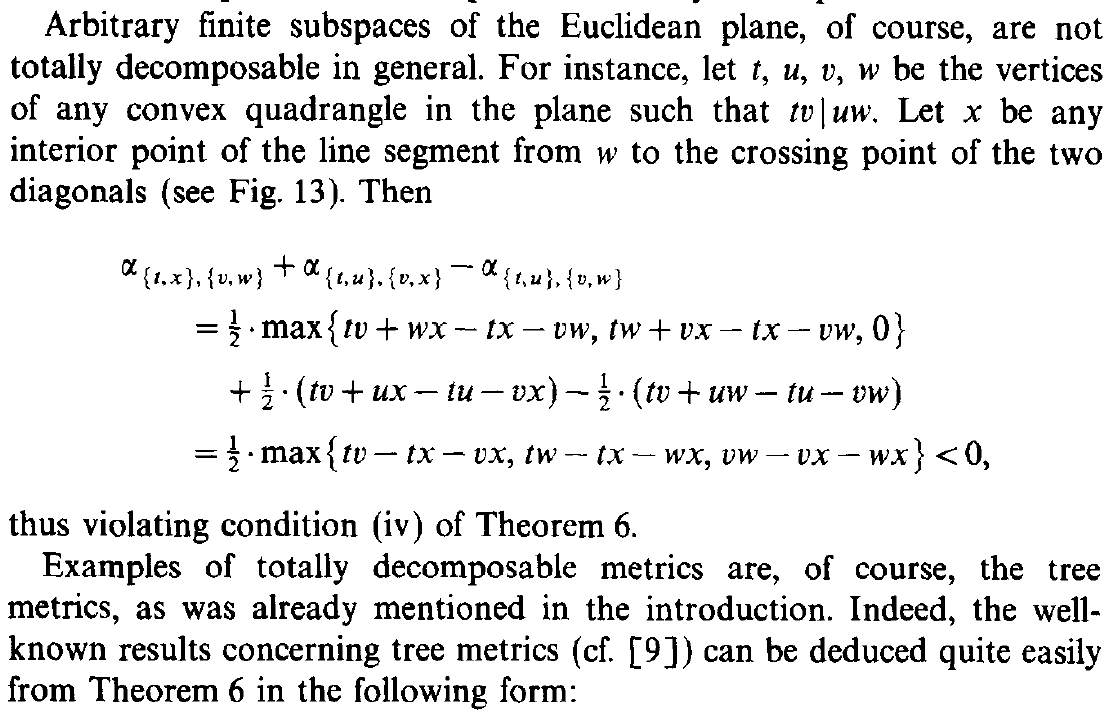
* check the identity for all quartets   
  (only those whose partial split have positive isolation index)
* if the identity holds, then check if
* if this is true, then is a -split and its isolation index is

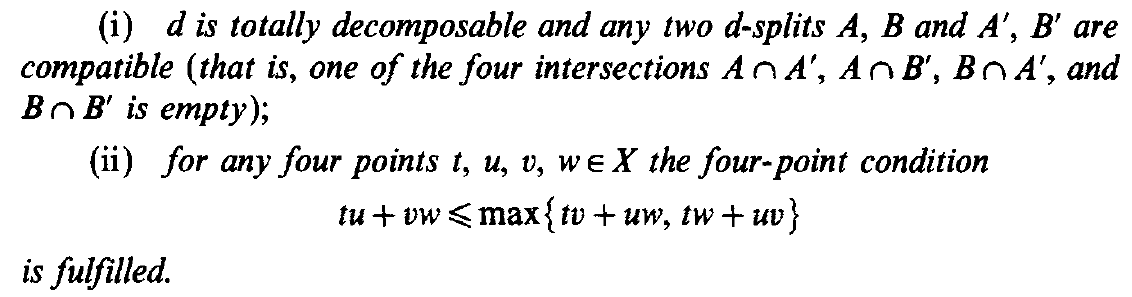
**Question**: Does it exist an algorithm for the general case?





[…]



**Def**. (compatible splits)

Given two splits and we say that they are **compatible** if  
 one of the following four intersections is empty

We say that a set of splits is **compatible**  
 if its splits are (pairwise) compatible.

**Oss**. Subsets of compatible sets are compatible.

**Oss**. A compatible set is weakly compatible.

In fact, if we call the sets of splits extending the three respective quartets,   
then weak compatibility is equivalent to saying that  
 at most two of them are non-empty,   
while compatibility is equivalent to saying that  
 at most one of them is non-empty  
 (the splits from two different sets are not compatible).

**Def**. (four-point condition)

We say that (or just ) satisfies the **four-point condition**  
 if for any four points it holds

or equivalently

**Dim**. It is clear.

Applying the first formula

**Oss**. If satisfies the four-point condition  
 and are such that , then

In fact, cannot be the maximum among the three (otherwise the isolation index would be zero).

**Prop**. If is a totally decomposable pseudo-metric, then

**Dim**. Let . Clearly a -split of is also a -split of , because by extending the isolation index does not increase (and so by restricting it does not decrease). So .

From **Teo**. 6, the isolation index of a -split of coincides with the sum of the indices of its extensions.   
In particular, there must be an extension that is also a -split of . Thus .

**Cor**. 7 Let be a pseudo-metric on .

Then is totally decomposable and any two -splits are compatible if and only if satisfies the four-point condition.

**Dim**. Let . Since compatibility is preserved by restriction, from the previous **Prop**. we get that is compatible for all subsets of .   
In particular is compatible.

By compatibility, at least two quartets cannot be -splits  
 – suppose WLOG and .

If , then

If , then

In both cases the four-point condition is satisfied.

Let . If , then for every

where we used the previous **Oss**. in the first line.

By **Teo**. 6 this proves that is totally decomposable.

Suppose to have two incompatible -splits and .  
Then we can suppose that exist such that

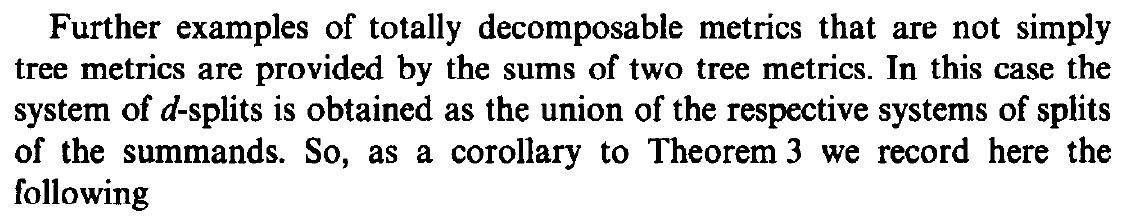
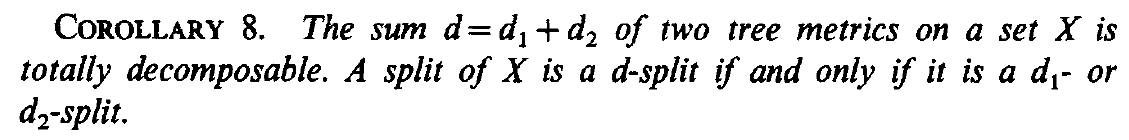
Since is totally decomposable, its restriction to is a linear combination of split metrics that splits 2-2 (that is they extend one of the quartets) and 3-1.   
But the latter contribute equally to the three distances.  
Since the set of -splits is weakly compatible and

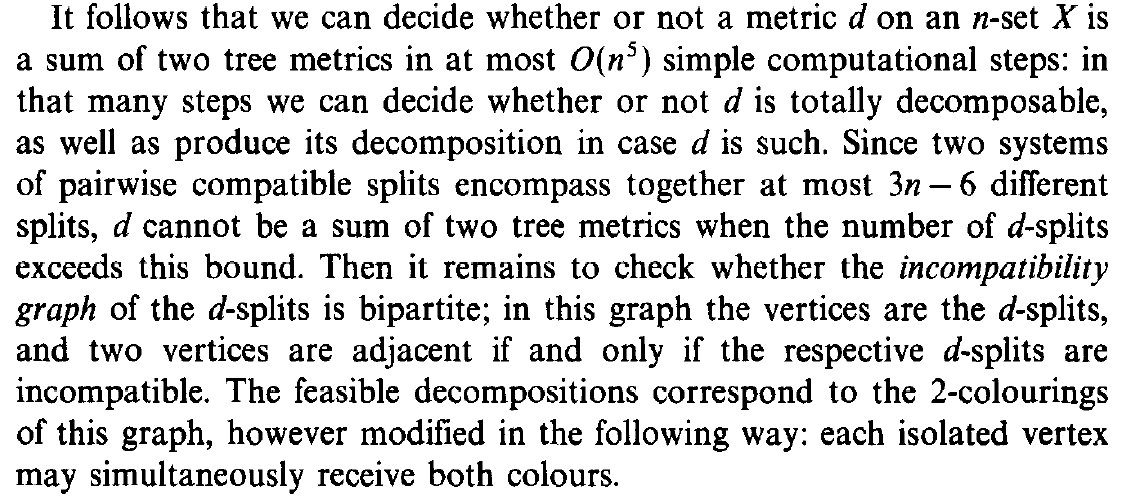
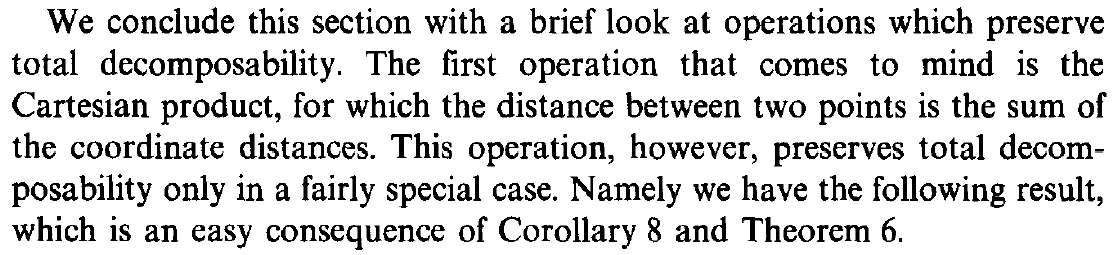
then does not have any -split extension.

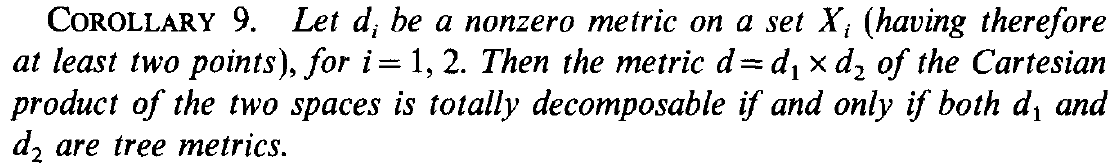
So the only different contributions come from split metrics that splits like and . As a consequence we have

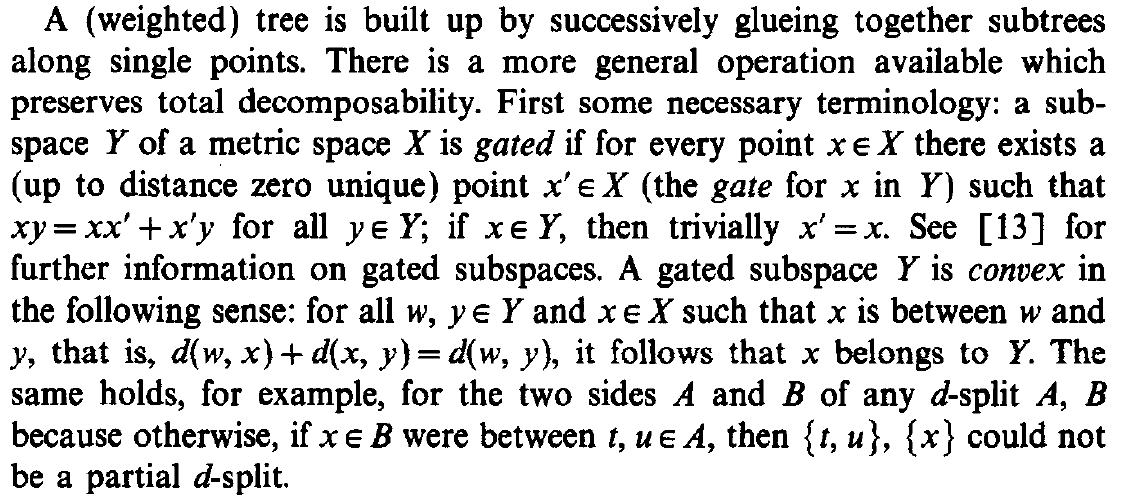
violating the four-point condition. ⭍

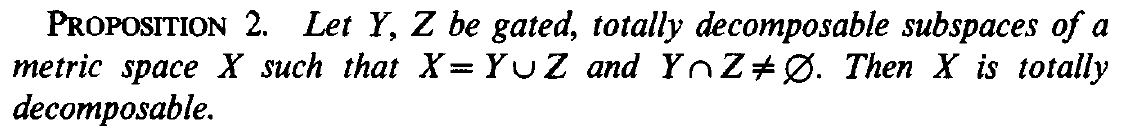
**Oss**. Tree metrics are totally decomposable pseudo-metrics  
 (since they satisfy the four-point condition).

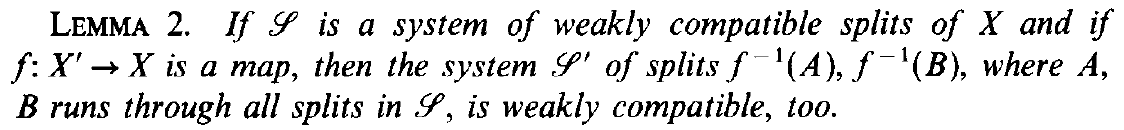
 

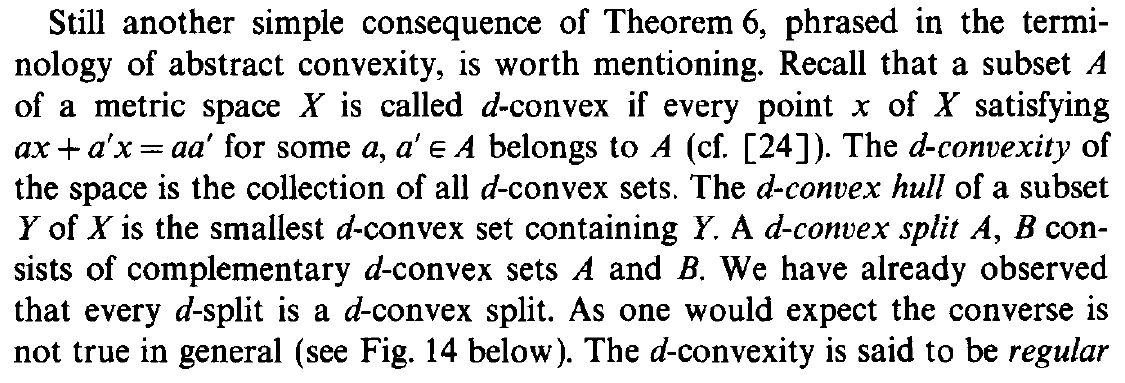
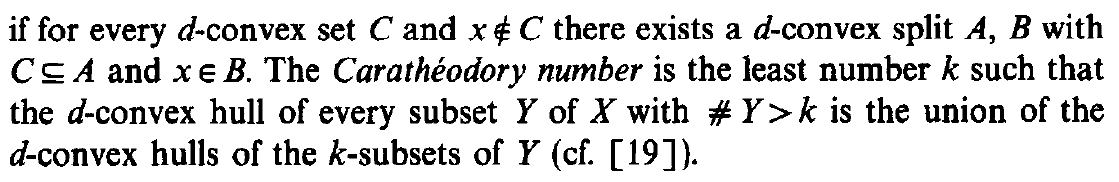
 

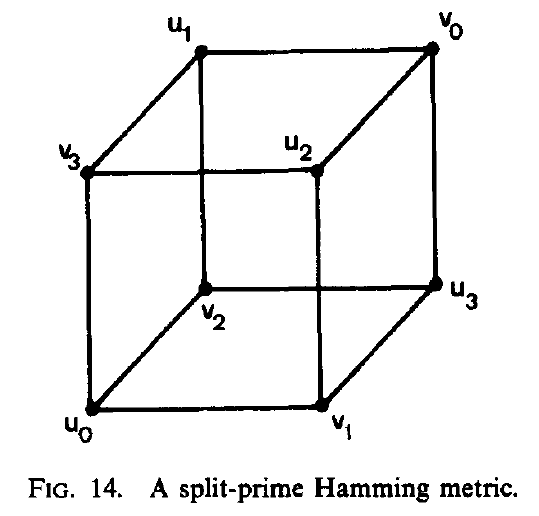


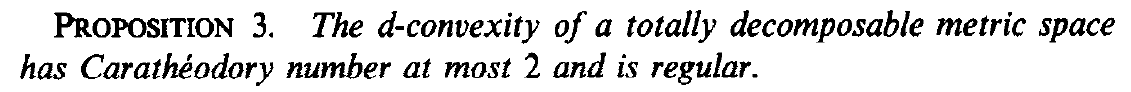


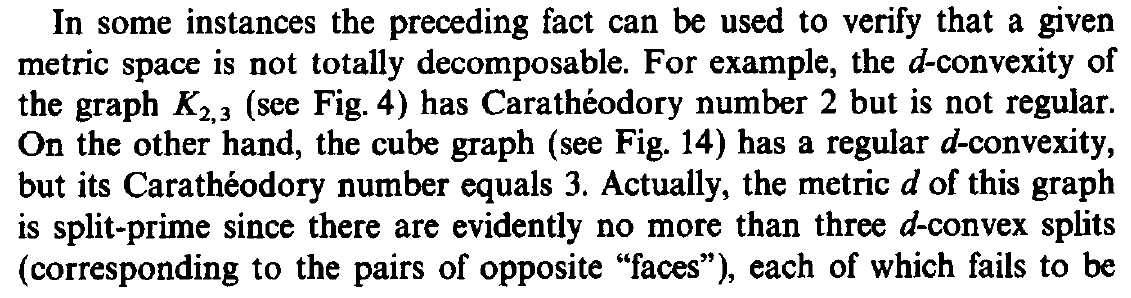
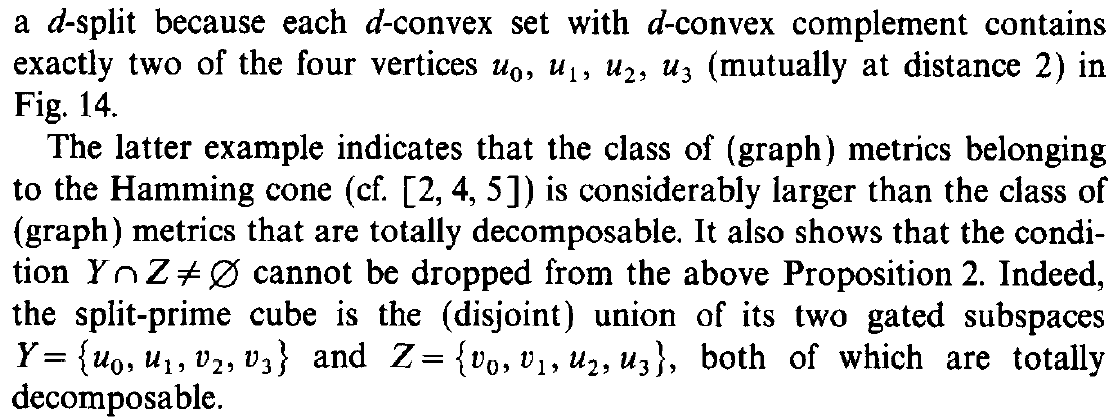




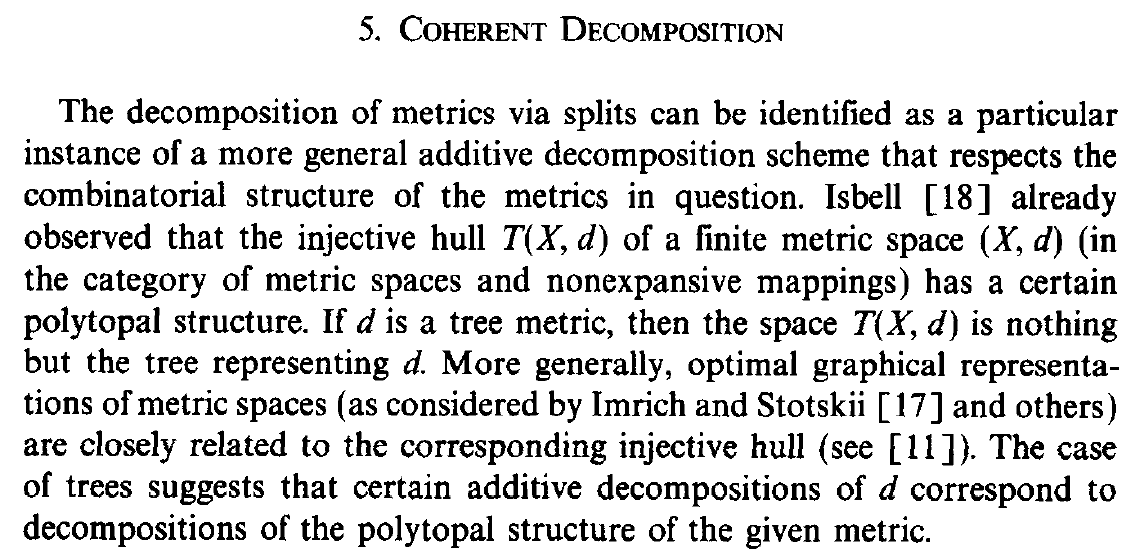
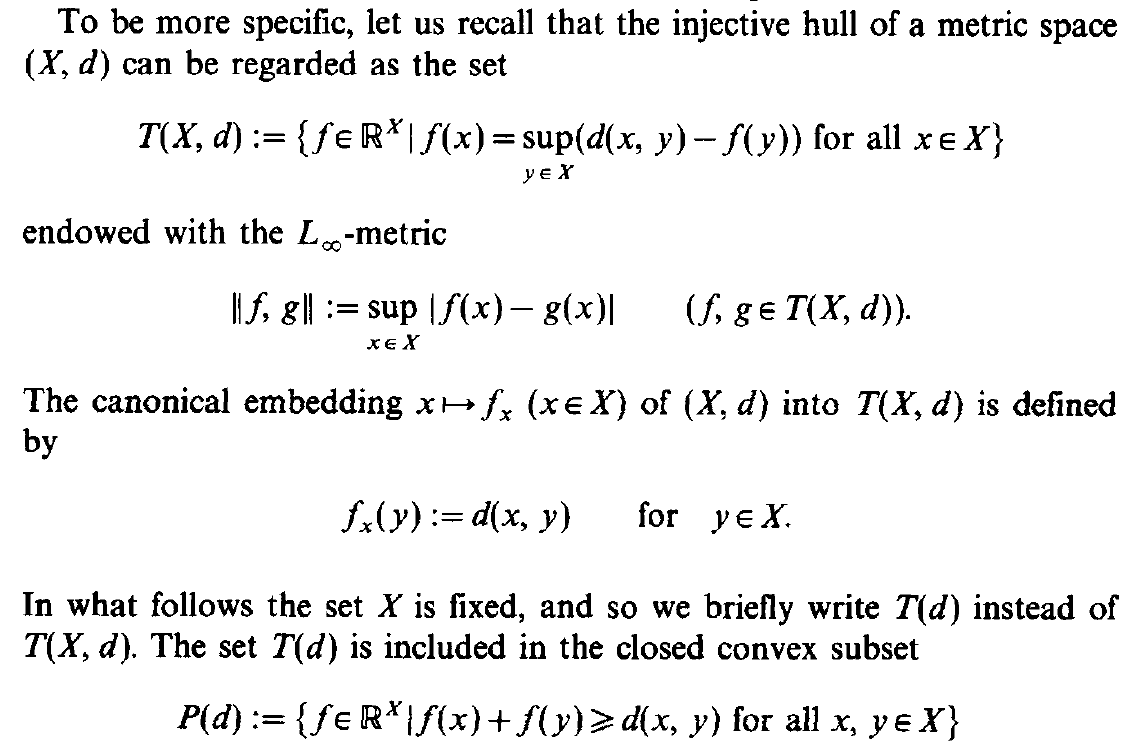
 

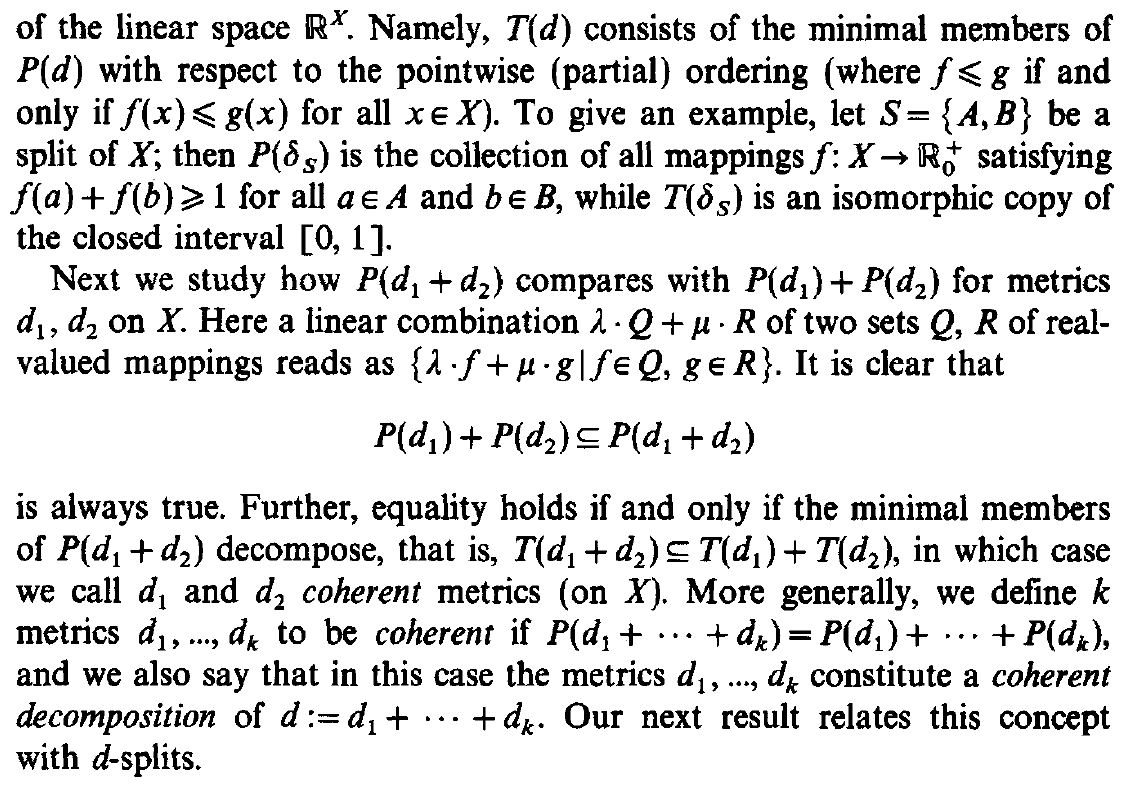


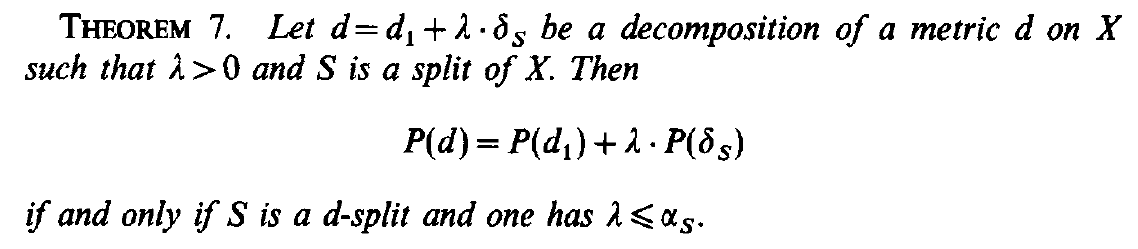


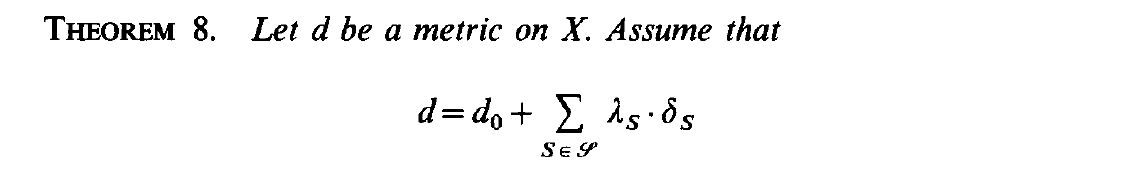
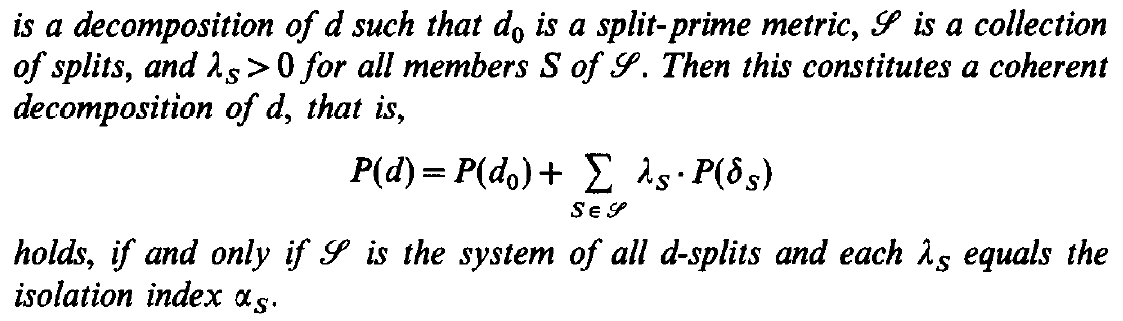
 

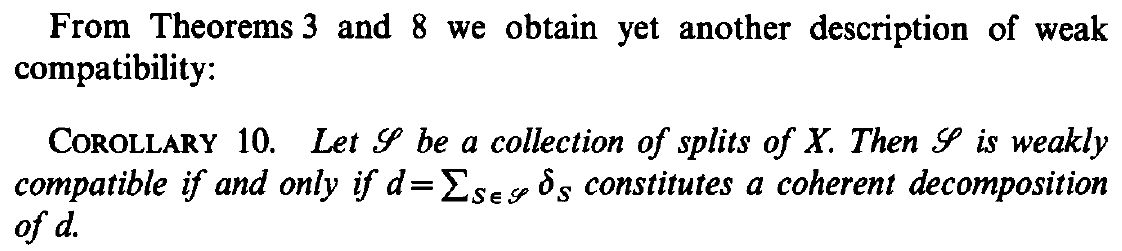
Chapter 5: Coherent Decomposition

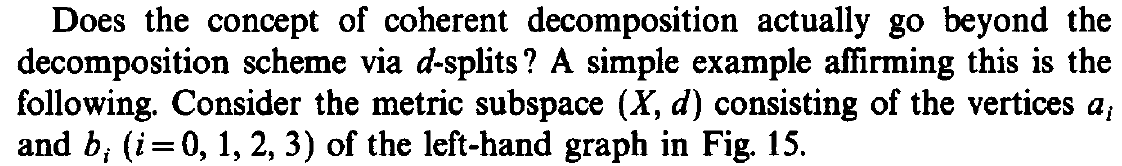
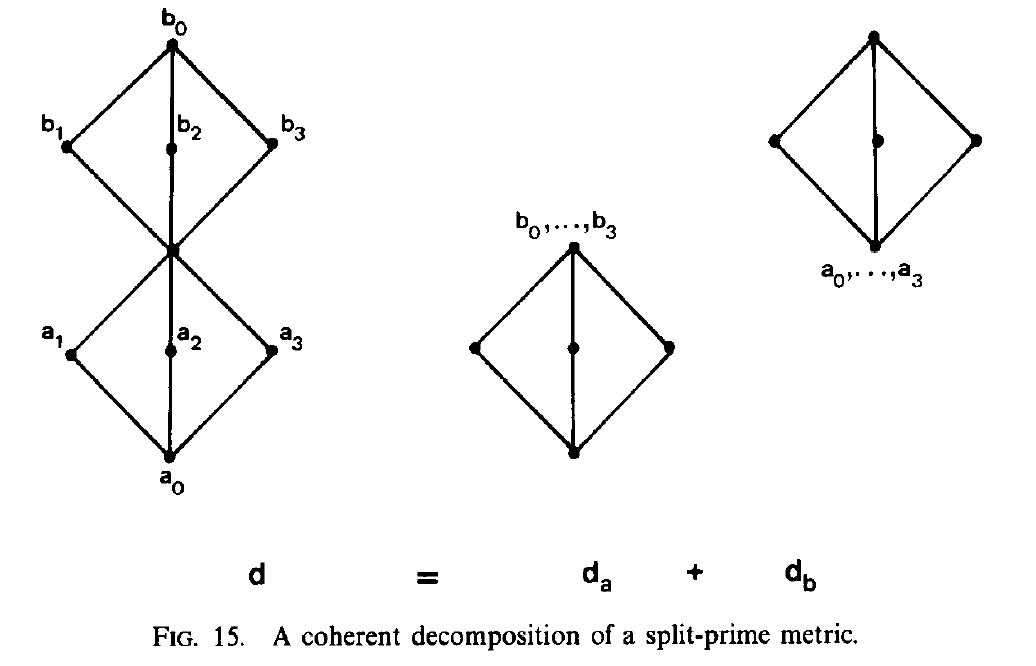
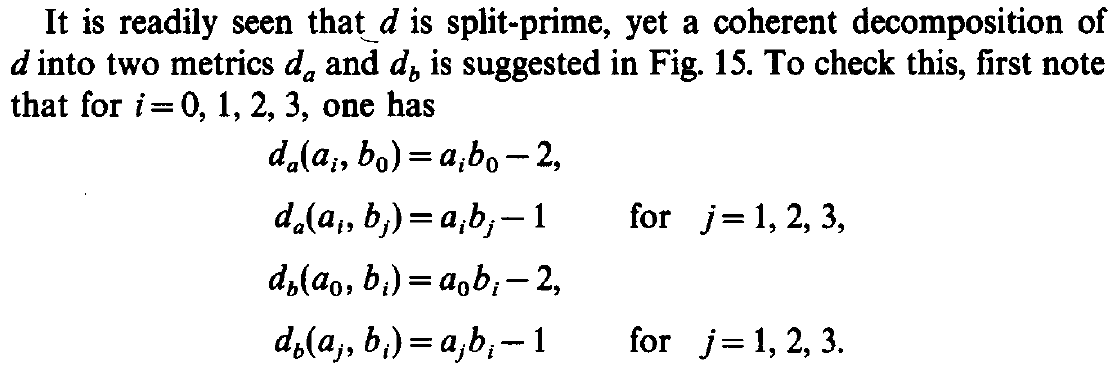
 

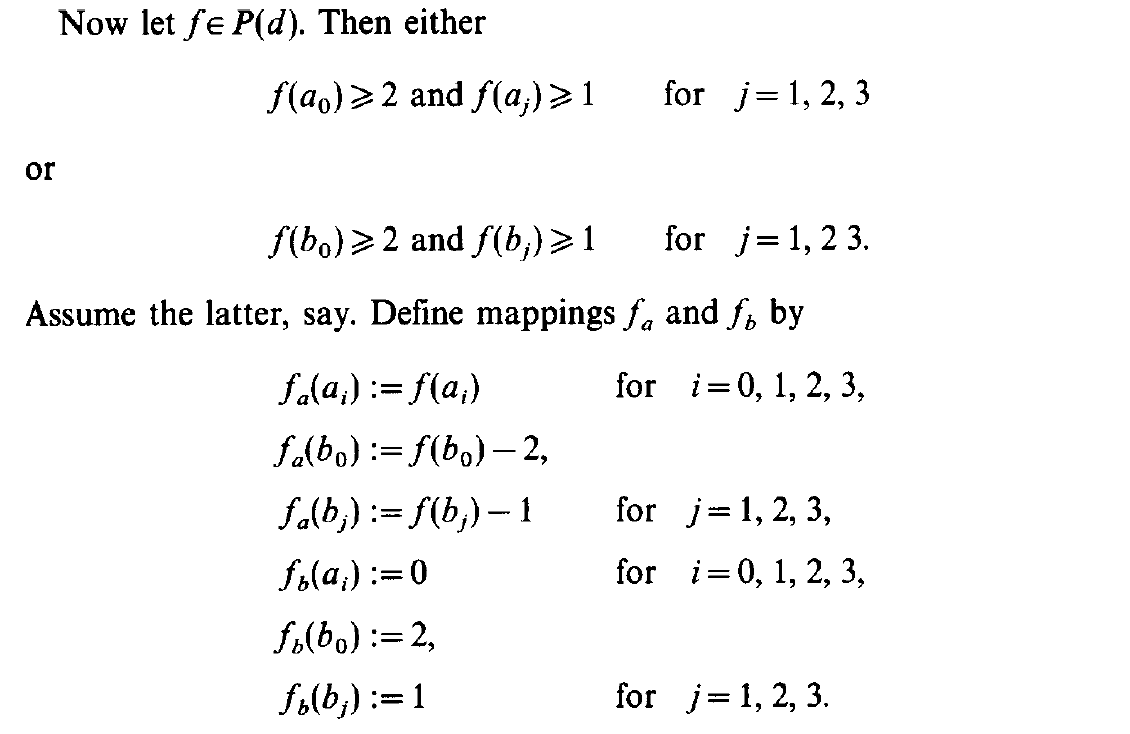
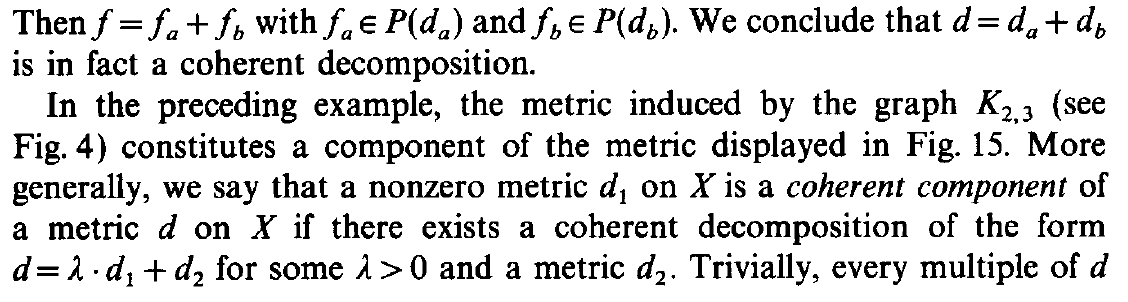
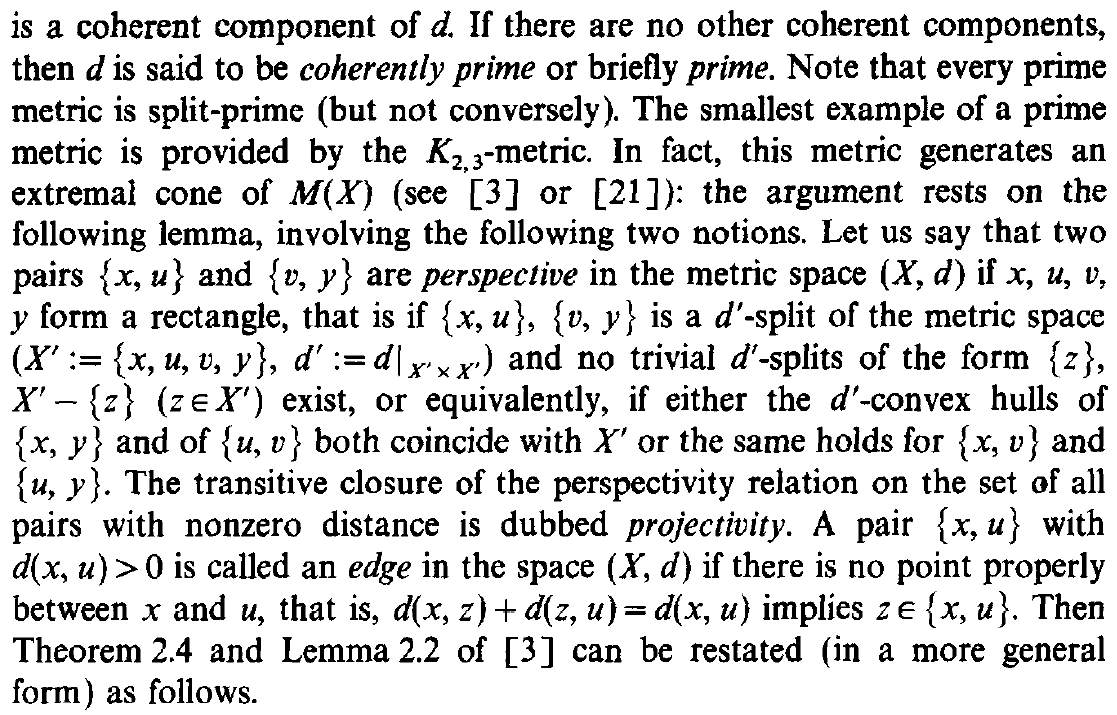
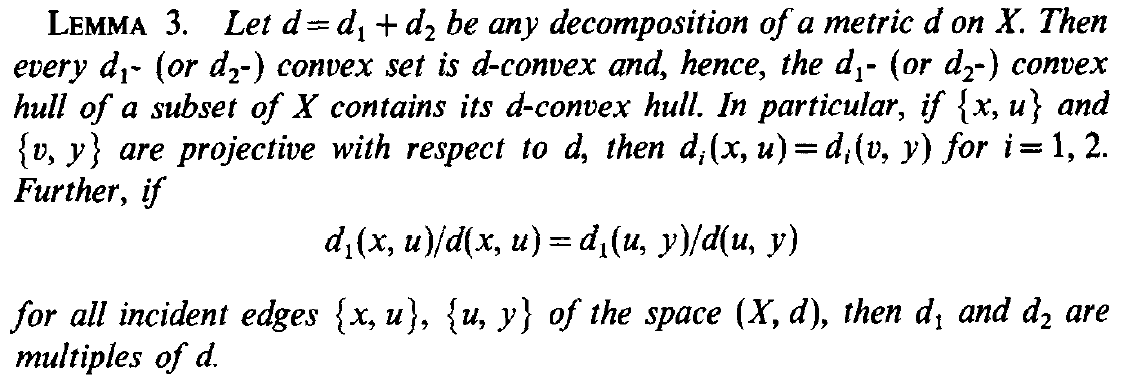
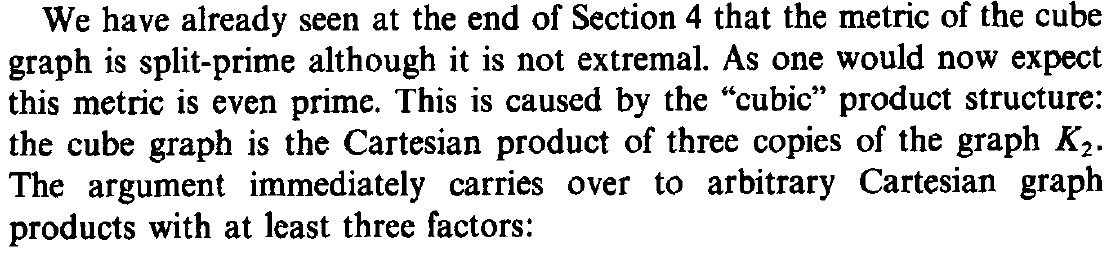
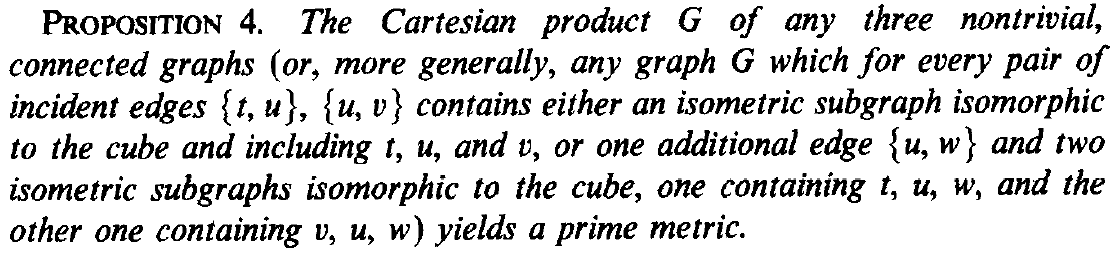


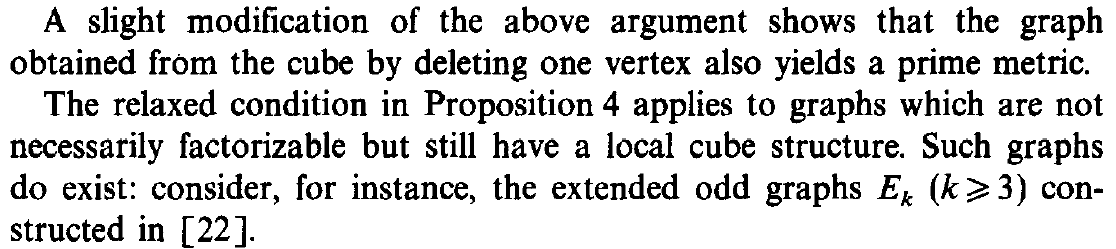
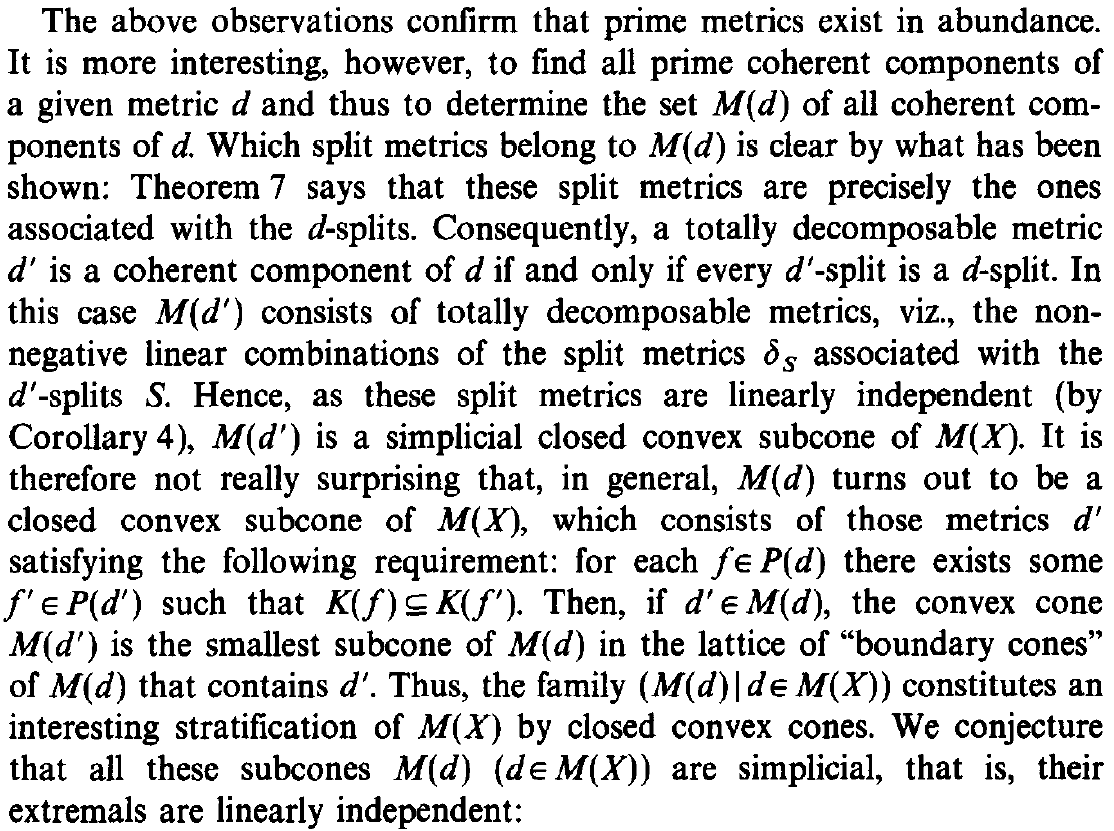
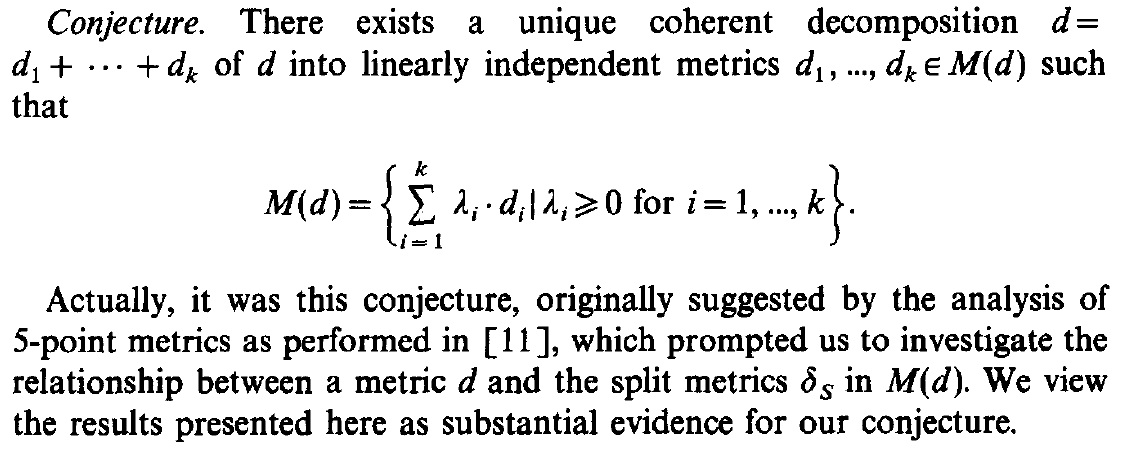


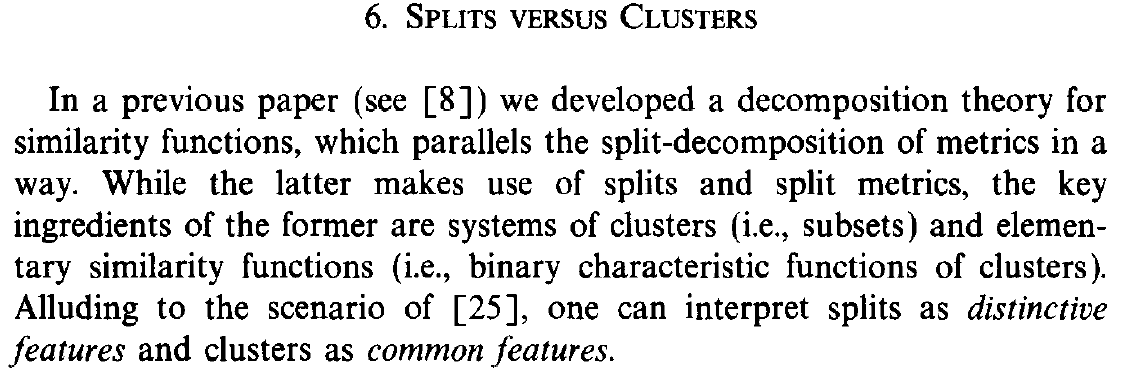
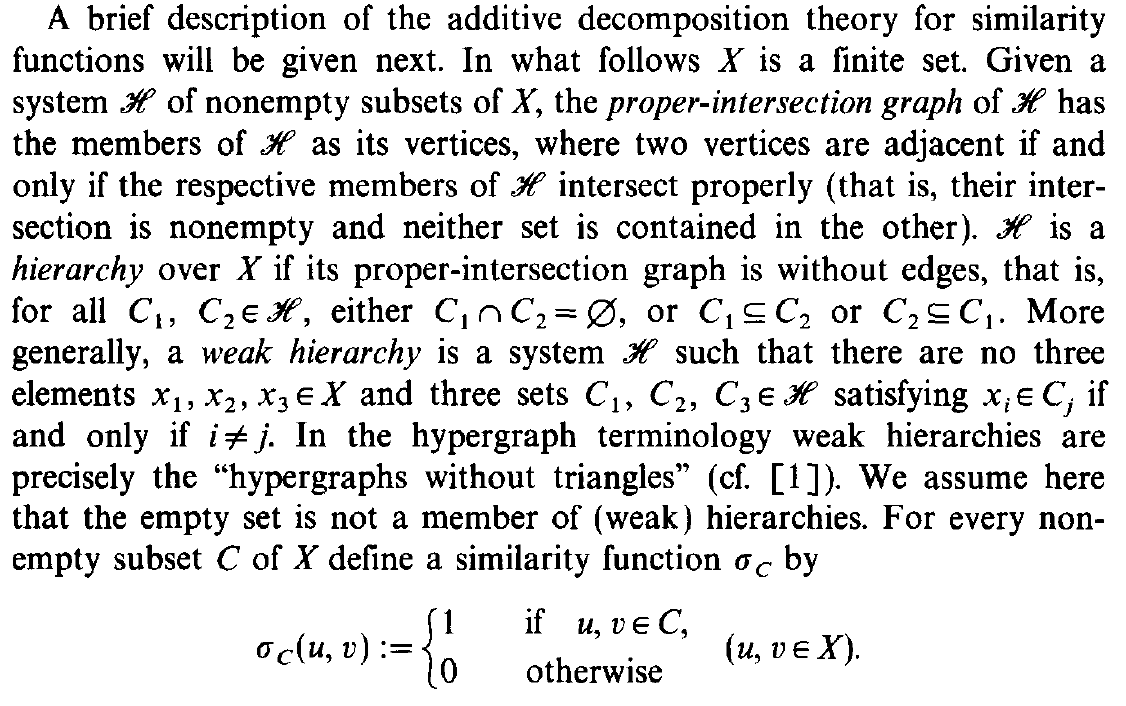


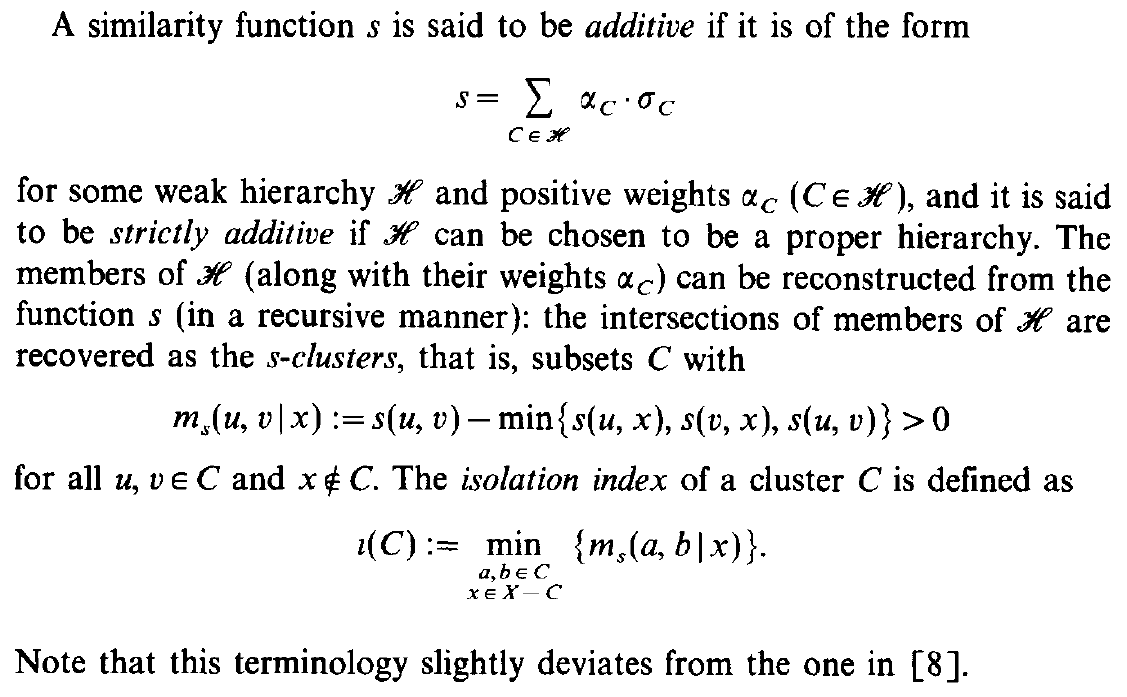
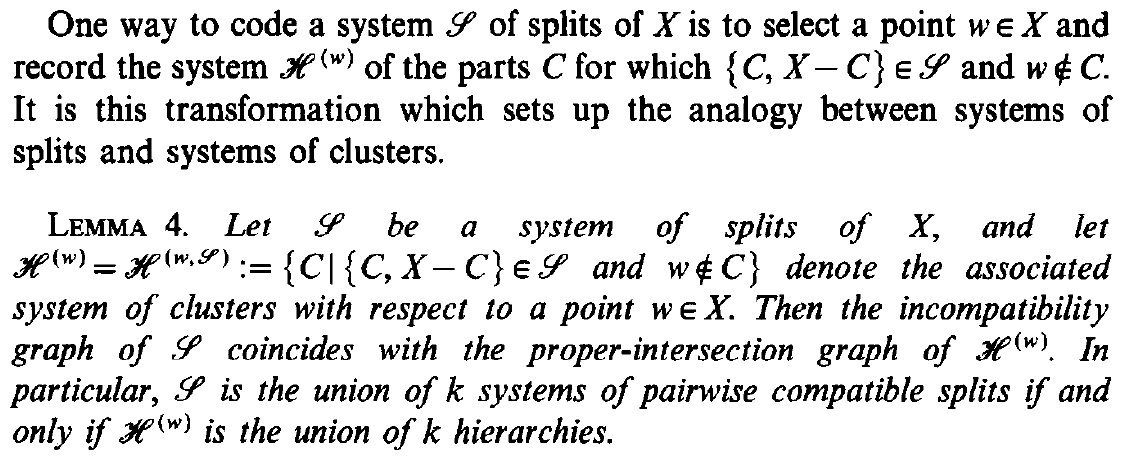
  

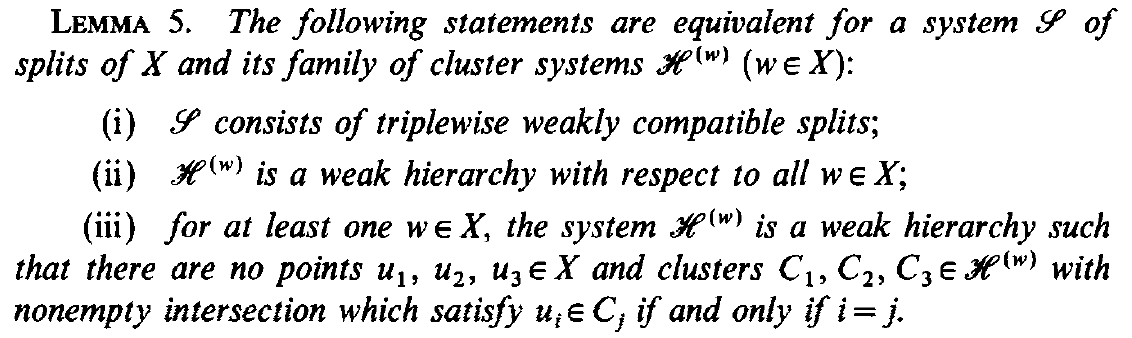
     

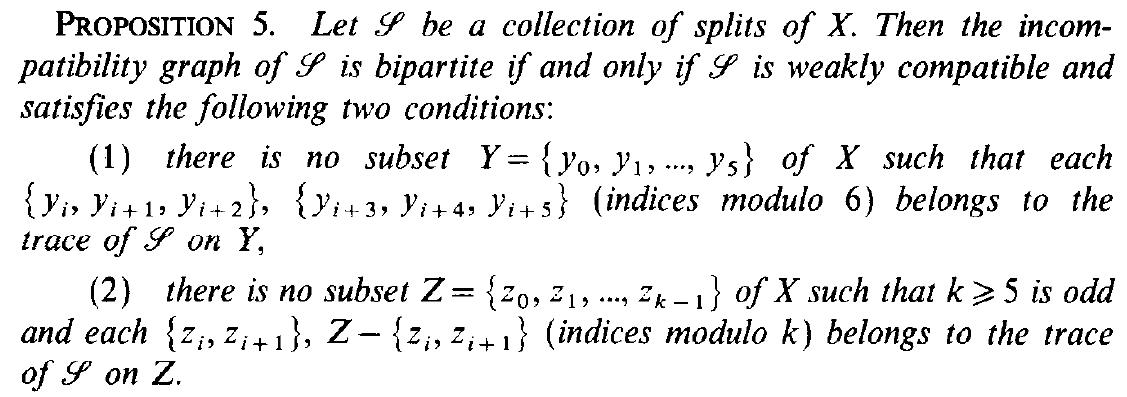
  

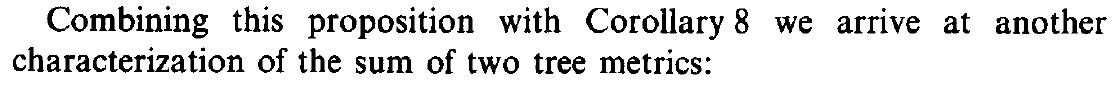
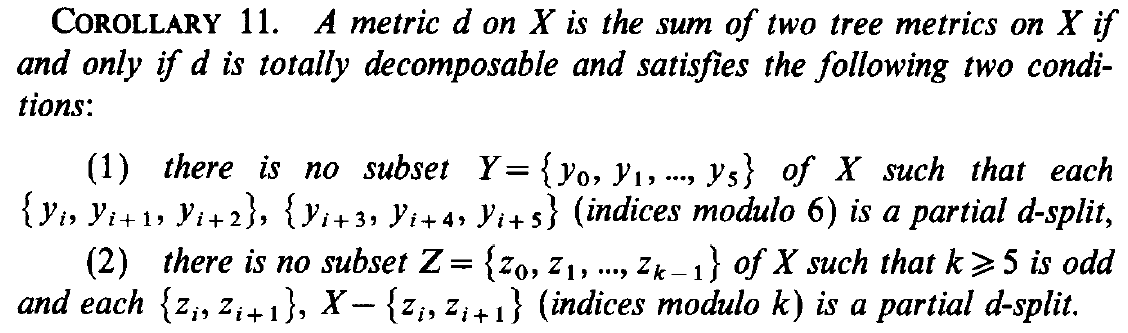
Chapter 6: Splits versus Clusters

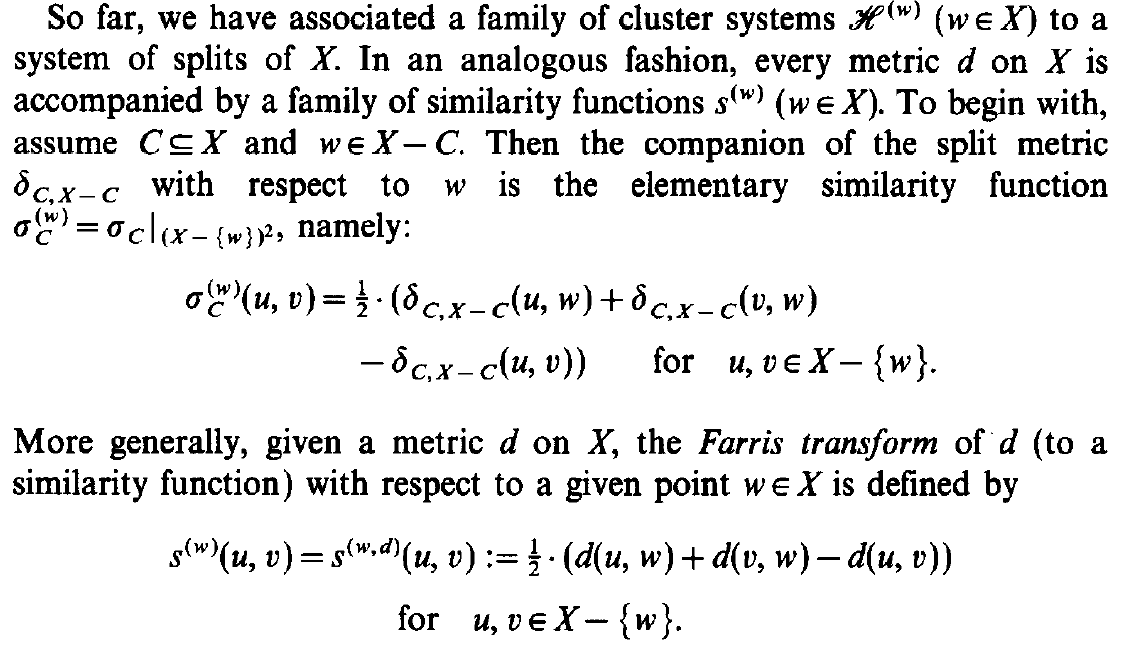
 

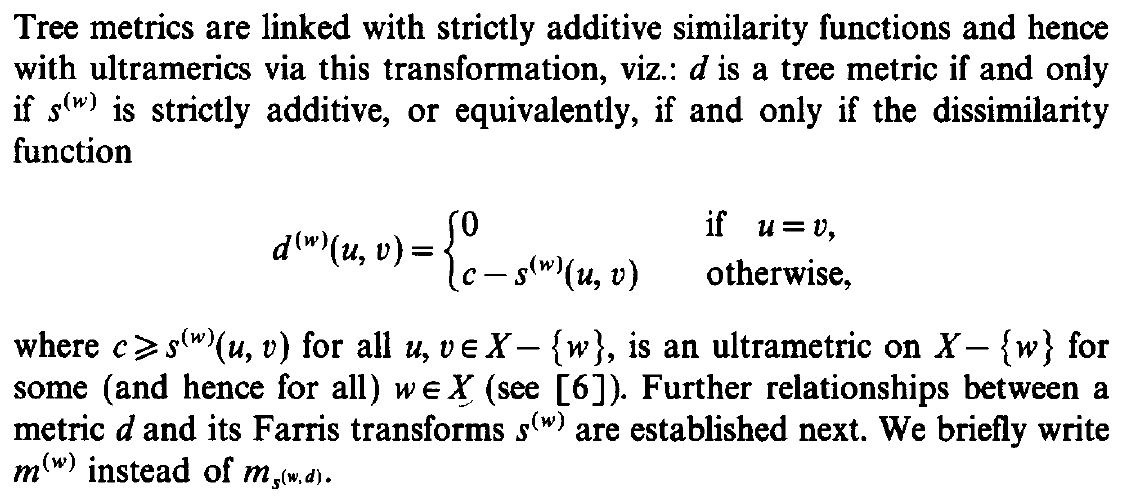
 

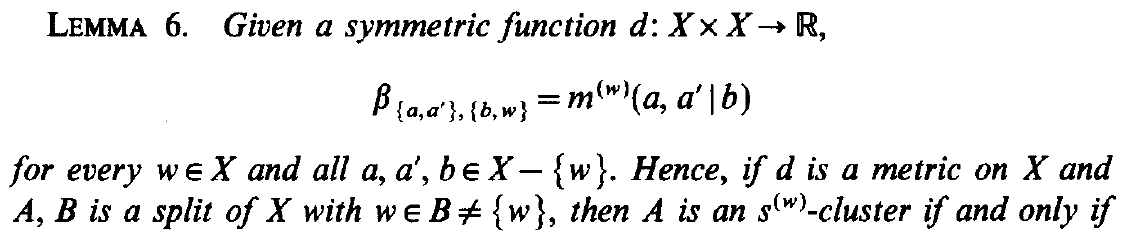
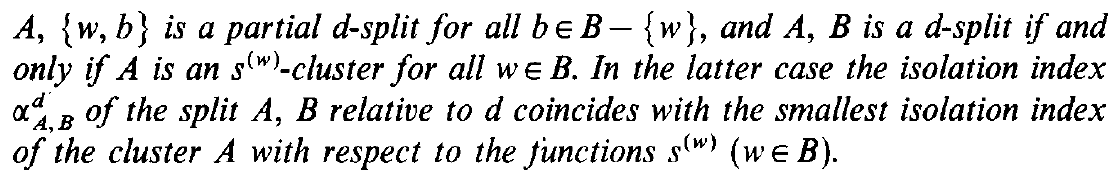


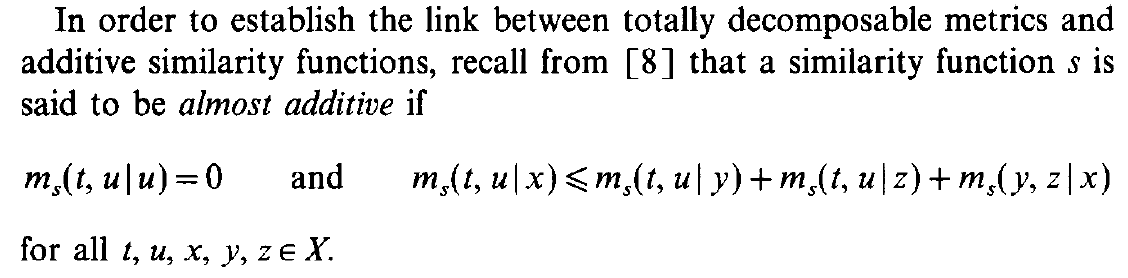


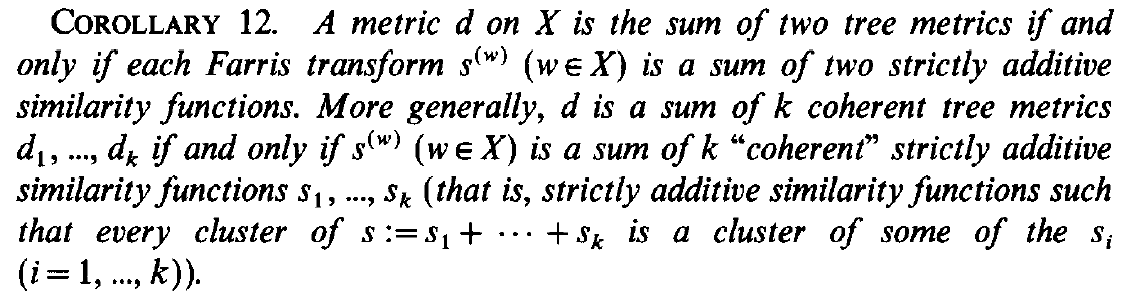


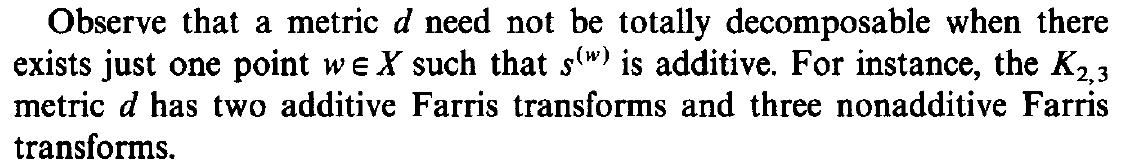
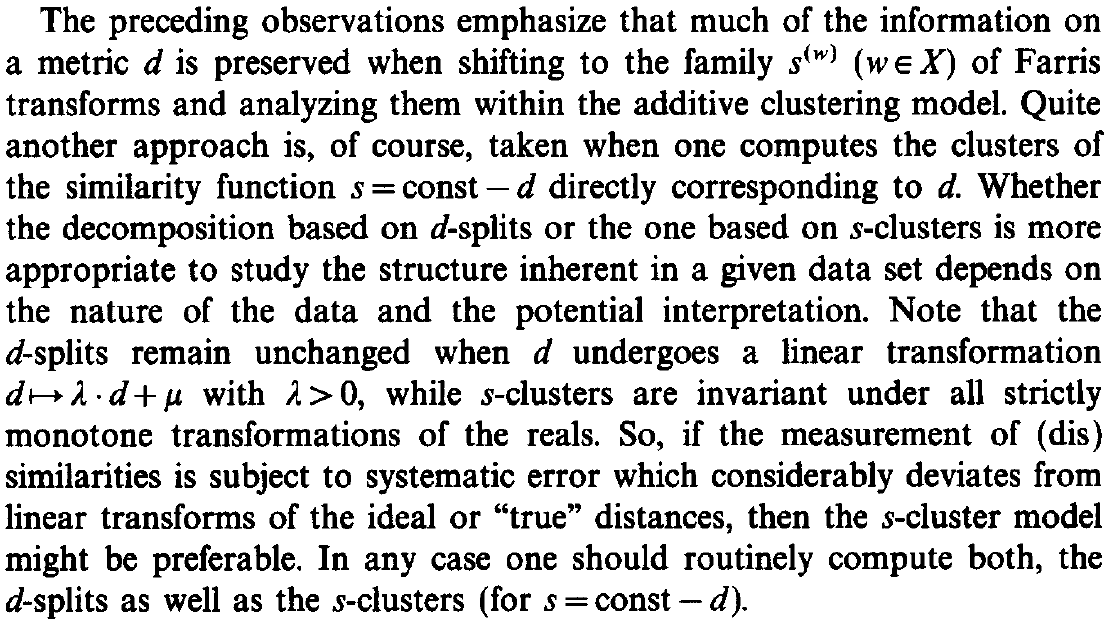


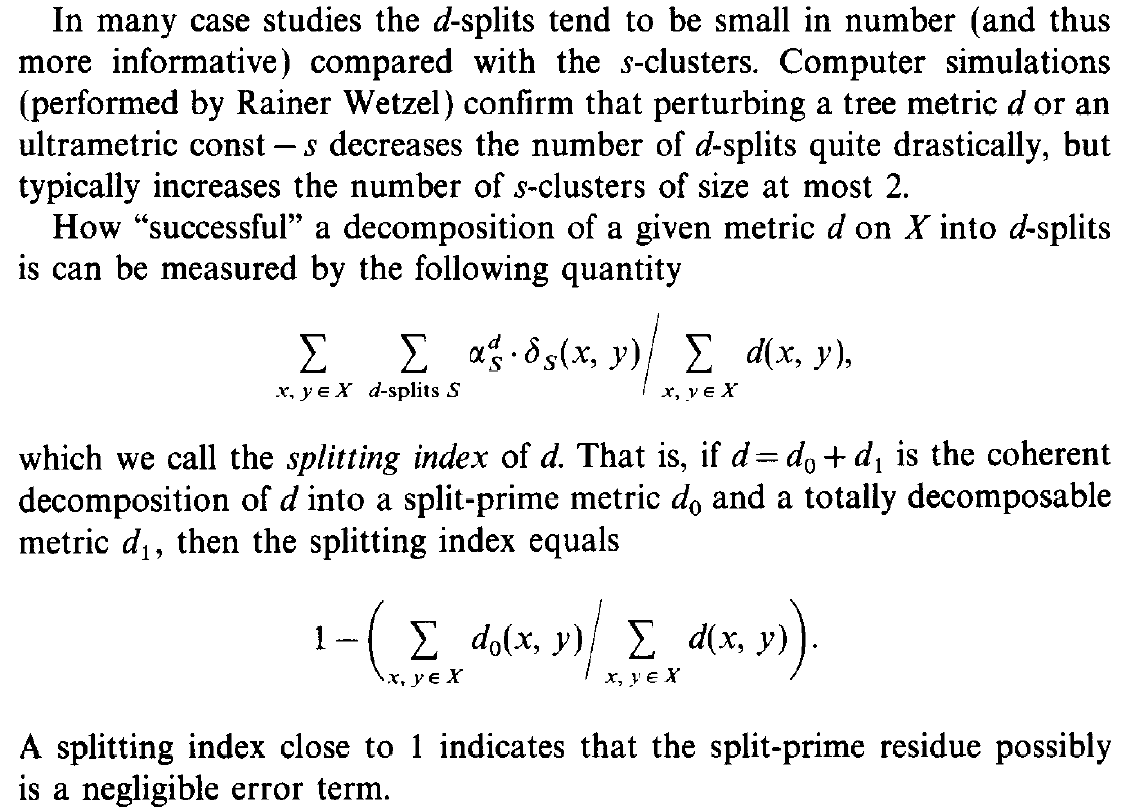
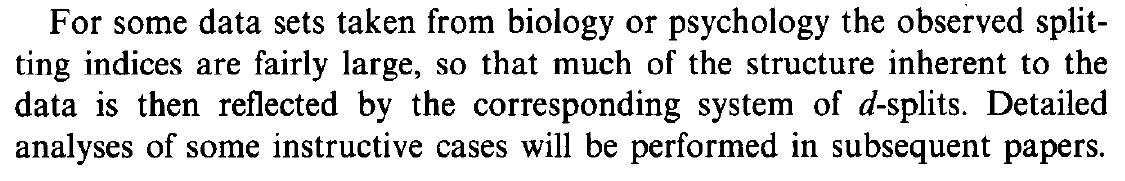
 







|  |  |
| --- | --- |
| **split** | **cluster** |
| incompatibility graph | proper-intersection graph |
| compatible splits | hierarchy |
| weakly compatible splits | weak hierarchy |
| metric | similarity function |
| tree metric | strictly additive similarity function |
| totally decomposable | additive / almost additive |
| linear transformations | strictly monotone transformations |

1. \* Here countable means finite or countably infinite (*al più numerabile*). [↑](#footnote-ref-1)
2. \* This is to be understood as every vector corresponds to an extreme ray  
    and no two vectors correspond to the same extreme ray. [↑](#footnote-ref-2)
3. \* For this reason, they may be also called extremal metrics. [↑](#footnote-ref-3)