Crypto Isomorphisms

in Phylogenetic Combinatorics

New Plymouth

January 05, 2003

X-Trees

Metrics over X

Split Systems over X

Quartet Systems over X

X-Trees

Definition

Given a finite set X, an X-Tree T is a finite tree $T = (V, E)$ with vertex set V and edge set E such that

- **X is the set of *leaves* of T,**
- **there is no vertex of degree 2 in V.**

Metrics over X

Metrics over X

Definition

A metric D over X is a map

$$D : X \times X \to \mathbf{R}$$

such that

- $D(a, b) = D(b, a)$,
- $D(a, c) \leq D(a, b) + D(b, c)$,
- **and** $D(a, a) = 0$

holds for all $a, b, c \in X$.

Split Systems over X

Split Systems over X

Definition

A *split system* $\mathcal{S}$ over X is a subset of the set $\mathcal{S}(X)$ consisting of all *splits of* X, i.e. of all pairs $\{A, B\}$ of subsets A, B of X with

- $A \cap B = \emptyset$,
- $\emptyset \neq A, B$,
- **and** $A \cap B = X$.

Quartet Systems over X

Quartet Systems over X

Definition

A quartet system $\mathcal{Q}$ over X is a subset of the set $\mathcal{Q}(X)$ consisting of all *2,2-splits of* X, i.e. of all pairs $\{A, B\}$ of subsets A, B of X with

- $A \cap B = \emptyset$ **and**
- $\#A = \#B = 2$.

Quartet Systems over X

Definition

A quartet system $\mathcal{Q}$ over X is a subset of the set $\mathcal{Q}(X)$ consisting of all *2,2-splits of* X, i.e. of all pairs $\{A, B\}$ of subsets A, B of X with

- $A \cap B = \emptyset$ **and**
- $\#A = \#B = 2$**.**

If $A = \{a, a'\}$ and $B = \{b, b'\}$, we also write $aa'|bb'$ instead of $\{\{a, a'\}, \{b, b'\}\}$.

Relation with experimental data

Relation with experimental data

Experimental data often provide *characters* for X, i.e. maps

$$\varphi : X \to C, \;\; \varphi' : X \to C', \ldots$$

from the set X in question into some other sets.

Relation with experimental data

Experimental data often provide *characters* for X, i.e. maps

$$\varphi : X \to C, \ \ \varphi' : X \to C', \dots$$

from the set X in question into some other sets.

Metrics:

Relation with experimental data

Experimental data often provide *characters* for X, i.e. maps

$$\varphi : X \to C,\ \ \varphi' : X \to C', \ldots$$

from the set X in question into some other sets.

Metrics:

Given a family $\varphi_i : X \to C_i$ $(i \in I)$ of such characters,

- **metrics D_i on C_i $(i \in I)$ give rise to a metric $D : X \times X \to \mathbf{R}$ defined by**

  $$D(a, b) := \sum_{i \in I} D_i(\varphi_i(a), \varphi_i(b)),$$

  for all $a, b \in X$.

Relation with experimental data

Experimental data often provide *characters* for X, i.e. maps

$$\varphi : X \to C, \;\; \varphi' : X \to C', \ldots$$

from the set X in question into some other sets.

Split systems:

Relation with experimental data

Experimental data often provide *characters* for X, i.e. maps

$$\varphi : X \to C, \ \ \varphi' : X \to C', \ldots$$

from the set X in question into some other sets.

Split systems:

Given a family $\varphi_i : X \to C_i$ $(i \in I)$ of such characters,

- **any family $(\mathcal{S}_i)_{i \in I}$ of split systems $\mathcal{S}_i$ over C_i gives rise to an induced split system**

  $$\mathcal{S}_i := \cup_{i \in I} \varphi_i^{-1}(\mathcal{S}_i)$$

  where $\varphi_i^{-1}(\mathcal{S}_i)$ is defined by

  $$\varphi_i^{-1}(\mathcal{S}_i) := \{\{A, B\} \in \mathcal{S}(X) \,|\, \{\varphi_i(A), \varphi_i(B)\} \leq \mathcal{S}_i\}$$

Here, we write $\{A, B\} \leq \mathcal{S}$ for two subsets A, B of a set C and a split system $\mathcal{S}$ of C if there exists a split $S = \{A', B'\}$ in $\mathcal{S}$ with $A \subseteq A$‘ and $B \subseteq B$‘.

Relation with experimental data

Quartet systems:

Relation with experimental data

Quartet systems:

Rate matrices $R(a,b) = (R(a,b)_{i,j\in I})$, one for each pair a, b from a family of aligned sequences

- **as considered in the context of $\log\det$ methods (note that the $\log\det$ distance of a and b is just, by definition, the *trace* of $R(a,b)$)**
- **and recently computed extensively for many sequence families by Alex Grossmann,**

give rise, for any *2,2*-split $ab|cd$ of distinct sequences a, b, c, d from such a family, rather useful *confidence values*

$$\mathbf{Conf}(ab|cd) :=$$

$$\langle R(a,b) \mid R(c,d)\rangle = \sum_{i,j} R(a,b)_{i,j} R(c,d)_{i,j}$$

for the *2,2*-split $ab|cd$ to represent the phylogenetically "correct" quartet tree from which quartet systems can easily be gathered.

Relation with experimental data

Quartet systems:

In addition, *any* tree-reconstruction method based on such data and applied to all 4-subsets Y of X (where basically all such methods are fast) provides interesting quartet systems.

Topics to be studied

Topics to be studied

How to *encode* X-trees by split systems, quartet systems, and metrics over X?

Topics to be studied

How to *encode* X-trees by split systems, quartet systems, and metrics over X?

***Consistency conditions*: Which metrics, split, or quartet systems over X are induced by — and thus encode — an X-tree?**

Topics to be studied

How to *encode* X-trees by split systems, quartet systems, and metrics over X?

***Consistency conditions*: Which metrics, split, or quartet systems over X are induced by — and thus encode — an X-tree?**

How to *recode* the distance, split, or quartet code of an X-tree in terms of its corresponding other codes? This works for larger classes of objects, too, and the resulting maps have *adjoint functor* interpretations.

Topics to be studied

How to *encode* X-trees by split systems, quartet systems, and metrics over X?

***Consistency conditions*: Which metrics, split, or quartet systems over X are induced by — and thus encode — an X-tree?**

How to *recode* the distance, split, or quartet code of an X-tree in terms of its corresponding other codes? This works for larger classes of objects, too, and the resulting maps have *adjoint functor* interpretations.

How to *decode* X-tree codes? Applied to inconsistent data, this can produce X*-diagrams* or X*-networks*.

Topics to be studied

How to *encode* X-trees by split systems, quartet systems, and metrics over X?

***Consistency conditions*: Which metrics, split, or quartet systems over X are induced by — and thus encode — an X-tree?**

How to *recode* the distance, split, or quartet code of an X-tree in terms of its corresponding other codes? This works for larger classes of objects, too, and the resulting maps have *adjoint functor* interpretations.

How to *decode* X-tree codes? Applied to inconsistent data, this can produce X*-diagrams* or X*-networks*.

How to *remove inconsistencies* in metrics, split, and quartet systems?

Topics to be studied

How to *encode* X-trees by split systems, quartet systems, and metrics over X?

***Consistency conditions*: Which metrics, split, or quartet systems over X are induced by — and thus encode — an X-tree?**

How to *recode* the distance, split, or quartet code of an X-tree in terms of its corresponding other codes? This works for larger classes of objects, too, and the resulting maps have *adjoint functor* interpretations.

How to *decode* X-tree codes? Applied to inconsistent data, this can produce X*-diagrams* or X*-networks.*

How to *remove inconsistencies* in metrics, split, and quartet systems?

Note that this question includes in particular the question of how to decide whether, given some split or quartet system, or some map defined on a subset of $X \times X$, when does there exist a consistent extension?

Topics to be studied

How to *encode* X-trees by split systems, quartet systems, and metrics over X?

***Consistency conditions*: Which metrics, split, or quartet systems over X are induced by — and thus encode — an X-tree?**

How to *recode* the distance, split, or quartet code of an X-tree in terms of its corresponding other codes? This works for larger classes of objects, too, and the resulting maps have *adjoint functor* interpretations.

How to *decode* X-tree codes? Applied to inconsistent data, this can produce X-*diagrams* or X-*networks*.

How to *remove inconsistencies* in metrics, split, and quartet systems?

***Weighted* versions.**

Topics to be studied

How to *encode* X-trees by split systems, quartet systems, and metrics over X?

***Consistency conditions*: Which metrics, split, or quartet systems over X are induced by — and thus encode — an X-tree?**

How to *recode* the distance, split, or quartet code of an X-tree in terms of its corresponding other codes? This works for larger classes of objects, too, and the resulting maps have *adjoint functor* interpretations.

How to *decode* X-tree codes? Applied to inconsistent data, this can produce *X-diagrams* or *X-networks.*

How to *remove* inconsistencies in metrics, split, and quartet systems?

***Weighted* versions.**

***Rooted* X-trees —the *affine* versus the *projective* approach to tree reconstruction.**

Topics to be studied

How to *encode* X-trees by split systems, quartet systems, and metrics over X?

***Consistency conditions*: Which metrics, split, or quartet systems over X are induced by — and thus encode — an X-tree?**

How to *recode* the distance, split, or quartet code of an X-tree in terms of its corresponding other codes? This works for larger classes of objects, too, and the resulting maps have *adjoint functor* interpretations.

How to *decode* X-tree codes? Applied to inconsistent data, this can produce X*-diagrams* or X*-networks*.

How to *remove* inconsistencies in metrics, split, and quartet systems?

***Weighted* versions.**

***Rooted* X-trees —the *affine* versus the *projective* approach to tree reconstruction.**

Restrictions to subsets of X.

How to encode an X-tree

$$T = (V, E)$$

by a *split system* over X?

How to encode an X-tree

$$T = (V, E)$$

by a *split system* over X?

Given any edge $e \in E$, the two subets A, B of X in the two connected components of

$$T_e := (V, E - \{e\})$$

form an X-split $S_e := \{A, B\}$.

How to encode an X-tree

$$T = (V, E)$$

by a ***split system*** over X?

Given any edge $e \in E$, the two subets A, B of X in the two connected components of

$$T_e := (V, E - \{e\})$$

form an X-split $S_e := \{A, B\}$.

One defines

$$\mathcal{S}_T := \{S_e \,|\, e \in E^*\}$$

where E^* denotes the set of internal edges of the tree T.

How to encode an X-tree

$$T = (V, E)$$

by a *quartet system* over X?

How to encode an X-tree

$$T = (V, E)$$

by a *quartet system* over X?

For any subset Y of X, let $T(Y)$ denote the Y-tree *induced* on Y by T, and put

$$\mathcal{Q}_T := \cup_{Y \in \binom{X}{4}} \mathcal{S}_{T(Y)}.$$

How to encode an X-tree

$$T = (V, E)$$

by a ***quartet system*** over X?

For any subset Y of X, let $T(Y)$ denote the Y-tree *induced* on Y by T, and put

$$\mathcal{Q}_T := \cup_{Y \in \binom{X}{4}} \mathcal{S}_{T(Y)}.$$

Note that $\mathcal{S}_{T(Y)}$ contains at most one split for every 4-subset Y of X, and that this split — if it exists at all — is always a *2,2*-split.

How to encode an X-tree

$$T = (V, E)$$

by a by *metric* defined on X?

How to encode an X-tree

$$T = (V, E)$$

by a *metric* defined on X?

One defines a metric

$$D = D_T : X \times X \to \mathbf{R}$$

by letting $D_T(a, b)$ denote the *length* of the (unique) path from a to b, for all a, b in X.

Theorem

Given two X-trees T and T', the following assertions are equivalent:

(i) T and T' are *canonically* isomorphic,

(ii) the splits systems $\mathcal{S}_T$ and $\mathcal{S}_{T'}$ coincide,

(iii) the quartet systems $\mathcal{Q}_T$ and $\mathcal{Q}_{T'}$ coincide,

(iv) the metrics D_T and $D_{T'}$ coincide.

Theorem

Given a split system $\mathcal{S} \subseteq \mathcal{S}(X)$ over X, there exists an X-tree T with

$$\mathcal{S}_T = \mathcal{S}$$

if and only if $\mathcal{S}$ is *compatible*, i.e. there exist subsets $A \in S$ and $A' \in S'$ with

$$A \cup A' = X$$

for any two splits S, S' in $\mathcal{S}$.

Theorem

Given a quartet system $\mathcal{Q} \subseteq \mathcal{Q}(X)$ over X, there exists an X-tree T with $\mathcal{Q}_T = \mathcal{Q}$ if and only if

- **$\mathcal{Q}$ is *thin*, i.e. one has**

  $$\#(\mathbf{Q} \cap \{ab|cd, ac|bd, ad|bc\}) \leq 1,$$

  for any four distinct elements a, b, c, d in X,

- **$\mathcal{Q}$ is *transitive*, i.e. one has**

  $$ab|cd, ab|de \in \mathbf{Q} \Rightarrow ab|ce \in \mathbf{Q}$$

  for any five distinct elements a, b, c, d, e in X,

- **and $\mathcal{Q}$ is *balanced*, i.e. one has**

  $$ab|cd, bc|de \in \mathbf{Q} \Rightarrow ab|de \in \mathbf{Q},$$

  also for any five distinct elements a, b, c, d, e in X.

Theorem

Given a metric

$$D : X \times X \to \mathbf{R} : (a,b) \mapsto ab$$

there exists an X**-tree** T **with**

$$D_T = D$$

if and only if

- D **satisfies the *four-point condition***

  $$ab + cd \leq \max(ac + bd, ad + bc)$$

  for all $a, b, c, d \in X$ **,**

- **there exists, for any** $a, a', b \in X$ **with** $a, a' \neq b$**, some** $b' \in X$

  $$ab + a'b' = 2 + aa' + bb'.$$

Remarkably, defining $D(ab:cd)$ **and** $D(ab:c)$ **for all** $a,b,c,d\in X$ **by**

$$D(ab:cd) := \frac{1}{2}\max\left(ac+bd, ad+bc\right) - (ab+cd)$$

and

$$D(ab:c) := \min\left(D(ab:cx) \mid x\in X, D(ab:cx)>0\right),$$

(with $D(ab:c) := \infty$ **in case** $D(ab:cx)\leq 0$ **holds for all** $x\in X$**, of course), the four-point condition is obviously equivalent with the condition that** $D(ab:cd)\geq 0$ **holds for all** $a,b,c,d\in X$ **while, assuming that the four-point condition holds, the second condition is equivalent with the assertion that**

$$D(ab:c) = 1$$

holds for all $a,b,c\in X$ **with** $a,b\neq c$**.**

And both conditions together imply that

$$1/2\big(ab+bc+ca\big)\in\mathbf{Z}$$

and, hence, in particular $ab\in\mathbf{Z}$ **holds for all** $a,b,c\in X$**.**

From split to quartet systems and back

From split to quartet systems and back

Given any split system $\mathcal{S}$ over X, one can associate with $\mathcal{S}$ the quartet system $\mathcal{Q} = \mathcal{Q}_{\mathcal{S}}$ over X defined by

$$\mathcal{Q}_{\mathcal{S}} := \{aa'|bb' \in \mathcal{Q}(X) \mid aa'|bb' \leq \mathcal{S}\}$$

or, equivalently, by

$$\mathcal{Q}_{\mathcal{S}} := \cup_{\{A,B\} \in \mathcal{S}} \; \mathcal{Q}_{\{A,B\}}$$

where $\mathcal{Q}_{\{A,B\}}$ denotes the set of all *2,2*-subsplits of a split $\{A, B\} \in \mathcal{S}(X)$, i.e. the set

$$\mathcal{Q}_{\{A,B\}} := \{aa'|bb' \in \mathcal{Q}(X) \,|\, a, a' \in A; b, b' \in B\}.$$

From split to quartet systems and back

Given any split system $\mathcal{S}$ over X, one can associate with $\mathcal{S}$ the quartet system $\mathcal{Q} = \mathcal{Q}_{\mathcal{S}}$ over X defined by

$$\mathcal{Q}_{\mathcal{S}} := \{aa'|bb' \in \mathcal{Q}(X) \mid aa'|bb' \leq \mathcal{S}\}.$$

And given any quartet system $\mathcal{Q}$ over X, one can associate with $\mathcal{Q}$ the split system $\mathcal{S} = \mathcal{S}_{\mathcal{Q}}$ over X defined by

$$\mathcal{S}_{\mathcal{Q}} := \{\{A, B\} \in \mathcal{S}(X) \mid \mathcal{Q}_{\{A,B\}} \subseteq \mathcal{Q}\}.$$

Clearly, one has

$$\mathcal{Q}_{\mathcal{S}_{\mathcal{T}}} = \mathcal{Q}_T \textbf{ and } \mathcal{S}_{\mathcal{Q}_{\mathcal{T}}} = \mathcal{S}_T$$

for every X-tree T.

From split to quartet systems and back

This sets up a *Galois correspondence* between split and quartet systems over X: Indeed, given any split system $\mathcal{S}$ and any quartet system $\mathcal{Q}$ over X, one has

$$\mathcal{Q}_{\mathcal{S}} \subseteq \mathcal{Q} \Leftrightarrow \mathcal{S} \subseteq \mathcal{S}_{\mathcal{Q}}.$$

Thus, putting $\mathcal{Q} := \mathcal{Q}_{\mathcal{S}}$ for some $\mathcal{S}$ as above, we see that

$$\mathcal{S} \subseteq \mathcal{S}_{\mathcal{Q}_{\mathcal{S}}}$$

holds for every split system $\mathcal{S}$, and putting $\mathcal{S} := \mathcal{S}_{\mathcal{Q}}$ for some $\mathcal{Q}$ as above, we see that, conversely,

$$\mathcal{Q}_{\mathcal{S}_{\mathcal{Q}}} \subseteq \mathcal{Q}$$

holds for every quartet system $\mathcal{Q}$.

Further, applying this last observation for some split system $\mathcal{S}$ and $\mathcal{Q}_0 := \mathcal{Q}_{\mathcal{S}}$, we get

$$\mathcal{Q}_{\mathcal{S}_{\mathcal{Q}_0}} \subseteq \mathcal{Q}_0 = \mathcal{Q}_{\mathcal{S}}$$

while $\mathcal{S} \subseteq \mathcal{S}_{\mathcal{Q}_{\mathcal{S}}} = \mathcal{S}_{\mathcal{Q}_0}$ implies

$$\mathcal{Q}_{\mathcal{S}} \subseteq \mathcal{Q}_{\mathcal{S}_{\mathcal{Q}_0}}.$$

From split to quartet systems and back

So, we have

$$\mathcal{Q} = \mathcal{Q}_{\mathcal{S}_{\mathcal{Q}}}$$

for some quartet system $\mathcal{Q}$ over X if and only $\mathcal{Q}$ is of the form $\mathcal{Q} = \mathcal{Q}_{\mathcal{S}}$ for some split system $\mathcal{S}$.

Similarly, one sees that

$$\mathcal{S} = \mathcal{S}_{\mathcal{Q}_{\mathcal{S}}}$$

holds for some split system $\mathcal{S}$ over X if and only $\mathcal{S}$ is of the form $\mathcal{S} = \mathcal{S}_{\mathcal{Q}}$ for some quartet system $\mathcal{Q}$.

It has been observed moreover in [A Relational Approach to Split decomposition] that the quartet systems $\mathcal{Q}$ over X with $\mathcal{Q} = \mathcal{Q}_{\mathcal{S}_{\mathcal{Q}}}$ can be characterized by a 7-point condition, and that $\{A_1 \cup A_2 \cup A_3, B_1 \cap B_2 \cap B_3\} \in \mathcal{S}_{\mathcal{Q}_{\mathcal{S}}}$ hold for any three splits $\{A_1, B_1\}, \{A_2, B_2\}, \{A_3, B_3\}$ in a split system $\mathcal{S}$ in case one has

$$A_1 \cup A_2 = A_2 \cup A_3 = A_3 \cup A_1.$$

From split to quartet systems and back

Moreover, many split systems including the compatible split systems and the *cyclic* split systems, i.e. the split systems $\mathcal{S} \subseteq \mathcal{S}(X)$ for which there exists a linear order $\leq$ on X such that for every split $S \in \mathcal{S}$ there exists some A in S with $A = \{x \in X \,|\, a \leq x \leq a'\}$ for some $a, a' \in A$, are known to satisfy the relation $\mathcal{S} = \mathcal{S}_{\mathcal{Q}_{\mathcal{S}}}$.

Note that the quartet systems $\mathcal{Q} = \mathcal{Q}_{\mathcal{S}}$ corresponding to maximal cyclic split systems can be characterized as follows: ...

From split systems to metrics and back

Given a split system $\mathcal{S}$ over X, one can associate with $\mathcal{S}$ the metric

$$D = D^{\mathcal{S}} : X \times X \to \mathbf{R} : (a, b) \mapsto ab^{\mathcal{S}}$$

over X that associates to any a, b in X the number $ab^{\mathcal{S}}$ of splits in the subset $\mathcal{S}(a|b)$ of $\mathcal{S}$ consisting of all splits in $\mathcal{S}$ that *separate* a from b, i.e. those S in $\mathcal{S}$ for which some $A \in S$ with $a \in A$ and $b \notin A$ exists, i.e. we define

$$\mathcal{S}(a|b) := \{\{A, B\} \in \mathcal{S} \mid a \in A, b \in B\}$$

and

$$D^{\mathcal{S}}(a, b) = ab^{\mathcal{S}} := \#\mathcal{S}(a|b).$$

From split systems to metrics and back

Conversely, given a metric $D : X \times X \to \mathbf{R} : (a, b) \mapsto ab$, one can associate with D the split system $\mathcal{S} = \mathcal{S}_D$ over X that contains exactly those splits

$$S = \{A, B\} \in \mathcal{S}(X)$$

for which

$$aa' + bb' < \min\left(ab + a'b', ab' + a'b\right)$$

holds for all $a, a' \in A$ and $b, b' \in B$.

From split systems to metrics and back

One can also associate with D the, generally larger, split system $\mathcal{S} = \mathcal{S}^D$ over X that contains exactly those splits

$$S = \{A, B\} \in \mathcal{S}(X)$$

for which

$$aa' + bb' < \max\left(ab + a'b', ab' + a'b\right)$$

or, equivalently,

$$D(aa' : bb') > 0$$

holds for all $a, a' \in A$ and $b, b' \in B$.

From split systems to metrics and back

Note that $\mathcal{S}_D = \mathcal{S}^D$ holds for all metrics D that satisfy the four-point condition.

Moreover, denoting the set of all *one-splits*, i.e. of all splits of X of the form $\big\{X - \{x\}, \{x\}\big\}$ for some x in X by $\mathcal{S}_1(X)$:

$$\mathcal{S}_1(X) := \Big\{\big\{X - \{x\}, \{x\}\big\} \,\Big|\, x \in X\Big\},$$

one has

$$\mathcal{S}_T \cap \mathcal{S}_1(X) = \emptyset$$

as well as

$$D^{\mathcal{S}_T \cup \mathcal{S}_1(X)} = D_T$$

and

$$\mathcal{S}_T \cup \mathcal{S}_1(X) = \mathcal{S}_{D_T} = \mathcal{S}^{D_T}$$

for every X-tree T.

From split systems to metrics and back

Note also that further interesting variations for associating a split system with a metric are possible.

From split systems to metrics and back

Theorem

Given any split system $\mathcal{S} \subseteq \mathcal{S}(X)$, the following assertions are equivalent:

- $\mathcal{S} = \mathcal{S}_D$ **holds for** $D := D^{\mathcal{S}}$**,**
- $\mathcal{S}$ **is of the form** $\mathcal{S}_D$ **for some metric** D **over** X**,**
- $\mathcal{S}$ **is compatible,**
- $\mathcal{S} - \mathcal{S}_1(X)$ **is of the form** $\mathcal{S}_T$ **for some** X**-tree** T**.**

Moreover, the metric $D^{\mathcal{S} \cup \mathcal{S}_1(X)}$ **associated with** $\mathcal{S} \cup \mathcal{S}_1(X)$ **is also of the form** D_T **for some** X**-tree** T **if these assertions hold.**

From split systems to metrics and back

Conversely, we have also the following

Theorem

Given any metric

$$D : X \times X \to \mathbf{R} : (a, b) \mapsto ab,$$

the following assertions are equivalent:

- $D = D^{\mathcal{S}}$ **holds for** $\mathcal{S} := \mathcal{S}_D$,
- D **is of the form** $D = D^{\mathcal{S}}$ **for some compatible split system** $\mathcal{S} \subseteq \mathcal{S}(X)$,
- **there exists a tree** $T = (V, E)$ **and a map** $\varphi : X \to V$ **with** $D(a, b) = D_T(\varphi(a), \varphi(b))$ **for all** $a, b \in X$
  (where $D_G(u, v)$ **denotes the length of the path from** u **to** v **in a connected graph** $G = (V, E)$**, for any two vertices** $u, v \in V$**),**
- **one has**

  $$D(ab : cd) \geq 0$$

  and $(ab + bc + ca)/2 \in \mathbf{Z}$ **for all** $a, b, c, d \in X$**, as well as**

  $$D(ab : c) = 1$$

  for all $a, b, c \in X$ **with** $a, b \neq c$**.**

From split systems to metrics and back

Note that, while $\mathcal{S}_D$ is always a compatible split system, the split system $\mathcal{S}^D$ is only *weakly compatible*, i.e. given any three splits

$$S_1 = \{A_1, B_1\}, S_2 = \{A_2, B_2\}, S_3 = \{A_3, B_3\},$$

one of the four sets

$$A_1 \cap A_2 \cap A_3,$$

$$A_1 \cap B_2 \cap B_3,$$

$$B_1 \cap A_2 \cap B_3,$$

or

$$B_1 \cap B_2 \cap A_3$$

is always empty, and that a split system $\mathcal{S}$ is weakly compatible if and only it is of the form $\mathcal{S} = \mathcal{S}^D$ for some metric D in which case one has $\mathcal{S} = \mathcal{S}^D$ in particular for $D := D^{\mathcal{S}}$.

From split systems to metrics and back

Next, we define

$$\mathcal{S}(A|B) := \{\{A', B'\} \in \mathcal{S} \mid A \subseteq A', B \subseteq B'\}$$

as well as

$$D_{\mathcal{S}}(A|B) := \#\mathcal{S}(A|B)$$

for any split system $\mathcal{S}$ and all subsets A, B of X, and

$$\mathcal{S}(a_1 \ldots a_n | b_1 \ldots b_m) := \mathcal{S}(\{a_1, \ldots, a_n\}|\{|b_1, \ldots, b_m\})$$

as well as

$$D_{\mathcal{S}}(a_1 \ldots a_n | b_1 \ldots b_m) := \#\mathcal{S}(a_1 \ldots a_n | b_1 \ldots b_m)$$

for all $a_1, \ldots, a_n, b_1, \ldots, b_m \in X$.

From split systems to metrics and back

Noting that

$$ab^{\mathcal{S}} = D_{\mathcal{S}}(a|b) =$$

$$D_{\mathcal{S}}(acd|b) + D_{\mathcal{S}}(ac|bd) + D_{\mathcal{S}}(ad|bc) + D_{\mathcal{S}}(a|cdb)$$

holds for any elements $a, b, c, d \in X$**, we get that**

$$ab^{\mathcal{S}} + a'b'^{\mathcal{S}} - aa'^{\mathcal{S}} - bb'^{\mathcal{S}} = 2\big(D_{\mathcal{S}}(aa'|bb') - D_{\mathcal{S}}(ab|a'b')\big)$$

and, therefore,

$$D^{\mathcal{S}}(aa' : bb') = D_{\mathcal{S}}(aa'|bb') - \min\big(D_{\mathcal{S}}(ab|a'b'), D_{\mathcal{S}}(ab'|a'b)\big)$$

holds for any four elements a, a', b, b' **in** X**.**

From split systems to metrics and back

Thus, taking into account that a split system $\mathcal{S}$ is weakly compatible if and only if at least one of the three sets $\mathcal{S}(aa'|bb'), \mathcal{S}(ab|a'b'), \mathcal{S}(ab'|a'b)$ is empty for any four elements a, a', b, b' in X, we see that

$$\max(D^{\mathcal{S}}(aa' : bb'), 0) = D_{\mathcal{S}}(aa'|bb')$$

holds for any four elements a, a', b, b' in X.

From split systems to metrics and back

Consequently, the metrics D with $D = D^{\mathcal{S}}$ for $\mathcal{S} := \mathcal{S}^D$ can be characterized as follows:

Theorem

Given any metric

$$D : X \times X \to \mathbf{R} : (a, b) \mapsto ab,$$

the following assertions are equivalent:

- $D = D^{\mathcal{S}}$ **holds for** $\mathcal{S} := \mathcal{S}^D$
- D **is of the form** $D = D^{\mathcal{S}}$ **for some weakly compatible split system** $\mathcal{S} \subseteq \mathcal{S}(X)$**,**
- **one has**
  $$\max(D(ab : cd), 0) \leq$$
  $$\max(D(ax : cd), 0) + \max(D(ab : xd), 0),$$
  $$D(ab : c) \in \{1, \infty\},$$
  and
  $$(ab + bc + ca)/2 \in \mathbf{Z}$$
  for all $a, b, c, d \in X$**.**

Please check wether this is really true?

From split systems to metrics and back

In consequence, associating the split system $\mathcal{S}_{D^{\mathcal{S}}}$ to a split system $\mathcal{S}$ defines an idempotent operator or *projection* that maps the set of all split systems over X onto the set of all compatible split systems over X, while associating $\mathcal{S}^{D^{\mathcal{S}}}$ to $\mathcal{S}$ defines a *projection* that maps the set of all split systems over X onto the set of all weakly compatible split systems over X.

Question: Given an arbitrary split system $\mathcal{S}$, how can we describe the compatible or weakly compatible split systems directly that result from applying either one of these projections?

From split systems to metrics and back

Associating with a single split $S = \{A, B\}$ in $\mathcal{S}(X)$ the metric $D^S := D^{\{S\}}$ that maps any x, y in X onto 0 in case $x, y \in A$ or $x, y \in B$ holds, and onto 1 else, does not only induce the map $\mathcal{S} \mapsto D^{\mathcal{S}}$ from split systems to metrics considered above.

It also induces a linear map

$$\lambda = \lambda_X : \mathbf{R}^{\mathcal{S}(X)} \to \mathcal{M}(X) : w \mapsto \sum_{S \in \mathcal{S}(X)} w(S)\, D^S$$

from the space $\mathbf{R}^{\mathcal{S}(X)}$ of all maps w from $\mathcal{S}(X)$ into $\mathbf{R}$ into the space $\mathcal{M}(X) = \mathcal{M}(X, \mathbf{R})$ of all symmetric maps D from $X \times X$ into $\mathbf{R}$ that vanish on the diagonal, as well as — by restriction — an additive map

$$\lambda\big|_{\mathbf{Z}} : \mathbf{Z}^{\mathcal{S}(X)} \to \mathcal{M}(X, \mathbf{Z})$$

from the $\mathbf{Z}$-lattice $\mathbf{Z}^{\mathcal{S}(X)}$ of all integer-valued maps w in $\mathbf{R}^{\mathcal{S}(X)}$ into the $\mathbf{Z}$-lattice $\mathcal{M}(X, \mathbf{Z})$ of all integer-valued maps in $\mathcal{M}(X, \mathbf{R})$.

From split systems to metrics and back

Remarkably,

- **the *support***

  $$\mathrm{supp}(r) := \{S \in \mathcal{S}(X) \,|\, r(S) \neq 0\}$$

  of every non-zero element r in the kernel of λ has cardinality at least 7 (for a simple proof, see next page),

- **defining, for each partition of X into four non-empty subsets A_1, A_2, A_3, A_4, the map $r_{A_1,A_2,A_3,A_4} \in \mathbf{Z}^{\mathcal{S}(X)}$ from $\mathcal{S}(X)$ into Z by**

  $$r_{A_1,A_2,A_3,A_4}(\{A_i, X - A_i\}) := 1$$

  for $i = 1, \ldots, 4$,

  $$r_{A_1,A_2,A_3,A_4}(\{A_i \cup A_j, X - (A_i \cup A_j)\}) := -1$$

  for $1 \leq i < j \leq 3$, and $r_{A_1,A_2,A_3,A_4}(S) := 0$ for every other split S in $\mathcal{S}(X)$, we get elements in this kernel whose support contains exactly 7 distinct splits,

- **every other element r in this kernel is an R-linear — and in case $r \in \mathbf{Z}^{\mathcal{S}(X)}$ even a Z-linear — combination of these specific elements.**

Please check whether this is really true.

From split systems to metrics and back

One can prove the fact that the support of every non-zero element r in the kernel of λ has cardinality at least 7 by induction on $n = \#X$ as follows: In case $n \leq 3$, the metrics of the form D^S with $S \in \mathcal{S}(X)$ are linearly independent, and in case $n = 4$, the only linear relation between such metrics is — up to a scalar — the relation

$$\sum_{S \in \mathcal{S}(X)} r_{A_1,A_2,A_3,A_4}(S)\, D^S = 0$$

with A_1, A_2, A_3, A_4 being the only partition — up to relabeling — of X into four distinct non-empyty subsets involving all 7 non-trivial splits of X.

Now assume that there are no elements r in the kernel of λ_Y with support of cardinality less than 7 for all proper subsets Y of X. And assume that $r \in \mathbf{R}^{\mathcal{S}(X)}$ is an element in the kernel of λ_Y with a support of cardinality at most 6.

Clearly, the relation

$$\sum_{S \in \mathcal{S}(X)} r(S)\, D^S = 0$$

and our inductive assuption applied to $Y := X - \{x\}$ for any $x \in X$ implies that for every split $S = \{A, B\}$ in $\mathrm{supp} r$ **with $x \in A \neq \{x\}$, we must have $0 \neq r(S) = -r(\{A - \{x\}, B \cup \{x\}\})$ and therefore $\{A - \{x\}, B \cup \{x\}\} \in \mathrm{supp} r$. However, this implies that $\mathrm{supp} r = \mathcal{S}(X)$ must hold in contradiction to $\#X > 3$ and $\#\mathrm{supp} r \leq 6$ because we can transform any split $S \in \mathcal{S}(X)$ into any one-split $\big\{a, X - \{a\}\big\}$ by removing, step by step, any element $x \neq a$ from the subset $A \in S$ with $a \in A$. So, assuming that there is at least one split in $\mathrm{supp} r$, $\mathrm{supp} r$ must contain every other split as well.**

More precisely, the above argument shows that there is — up to scaling — exactly one r in $\mathbf{R}^{\mathcal{S}(X)}$ that "restricts" to 0 on every proper subset Y of X in case $\#X$ is even, and none if $\#X$ is odd.

From split systems to metrics and back

It follows in particular that we can view the pair $(\mathbf{Z}^{\mathcal{S}(X)}, \mathrm{ke}(\lambda\big|_{\mathbf{Z}}))$ as a 3-*error correcting code* over Z.

Reducing modulo 2, this code is in fact well known as the ??? code with parameters ???.

From split systems to metrics and back

Furthermore, it is easy to check that, given any split system $\mathcal{S}$ over X, one must have

$$1/2\big(ab^{\mathcal{S}} + bc^{\mathcal{S}} + ca^{\mathcal{S}}\big) \in \mathbf{Z}$$

as before for any three elements because the sum $ab^{\mathcal{S}} + bc^{\mathcal{S}} + ca^{\mathcal{S}}$ counts each split in

$$\mathcal{S}(a|bc) \cup \mathcal{S}(b|ca) \cup \mathcal{S}(c|ab)$$

exactly twice.

However, it is not so easy to replace the condition

$$\max(D(ab:cd),0) \leq$$
$$\max(D(ax:cd),0) + \max(D(ab:xd),0),$$

above by one that, together with appropriate integrality conditions of the type considered above, is necessary and sufficient for D to be of the form $D = D^{\mathcal{S}}$ for some arbitrary split system $\mathcal{S}$ over X.

Please check what exactly is known or can be deduced about this.

From quartet systems to metrics and back

Given any quartet system $\mathcal{Q}$ over X and any two distinct elements $a, b \in X$, one can associate with $\mathcal{Q}$ and a, b the graph

$$G = G_{\mathcal{Q}|a,b} := (X, E_{\mathcal{Q}|a,b})$$

where $E_{\mathcal{Q};a,b}$ is defined by

$$E_{\mathcal{Q}|a,b} := \Big\{ \{c, d\} \in \binom{X - \{a, b\}}{2} \mid ab|cd \in \mathcal{Q} \Big\},$$

and then define a map

$$D = D_{\mathcal{Q}} : X \times X \to \mathbf{R}$$

by defining $D_{\mathcal{Q}}(a, b)$ to the number of connected components of $G_{\mathcal{Q}|a,b}$ diminished by 1.

Note that, in general, this need not be a metric.

From quartet systems to metrics and back

Conversely, given any metric

$$D : X \times X \to \mathbf{R},$$

one can associate with D **the quartet system** $\mathcal{Q} = \mathcal{Q}_D$ **over** X **consisting of all those** $ab|cd$ **in** $\mathcal{Q}(X)$ **for which**

$$D(a,b) + D(c,d) <$$

$$\min\big(D(a,c) + D(b,d), D(a,d) + D(b,c)\big)$$

or, alternatively, the quartet system $\mathcal{Q} = \mathcal{Q}^D$ **over** X **consisting of all those** $ab|cd$ **in** $\mathcal{Q}(X)$ **for which**

$$D(a,b) + D(c,d) <$$

$$\max\big(D(a,c) + D(b,d), D(a,d) + D(b,c)\big)$$

holds.

Also here, further interesting variations are possible.

Topics to be studied

How to *encode* X-trees by split systems, quartet systems, and metrics over X?

***Consistency conditions*: Which metrics, split, or quartet systems over X are induced by — and thus encode — an X-tree?**

How to *recode* the distance, split, or quartet code of an X-tree in terms of its corresponding other codes? This works for larger classes of objects, too, and the resulting maps have *adjoint functor* interpretations.

How to *decode* X-tree codes? Applied to inconsistent data, this can produce *X-diagrams* or *X-networks*.

How to *remove* inconsistencies in metrics, split, and quartet systems?

***Weighted* versions.**

***Rooted* X-trees —the *affine* versus the *projective* approach to tree reconstruction.**

Restrictions to subsets of X.

How to decode X-tree codes?

How to remove inconsistencies?

Weighted versions

Rooted X-trees

Restrictions to subsets of X