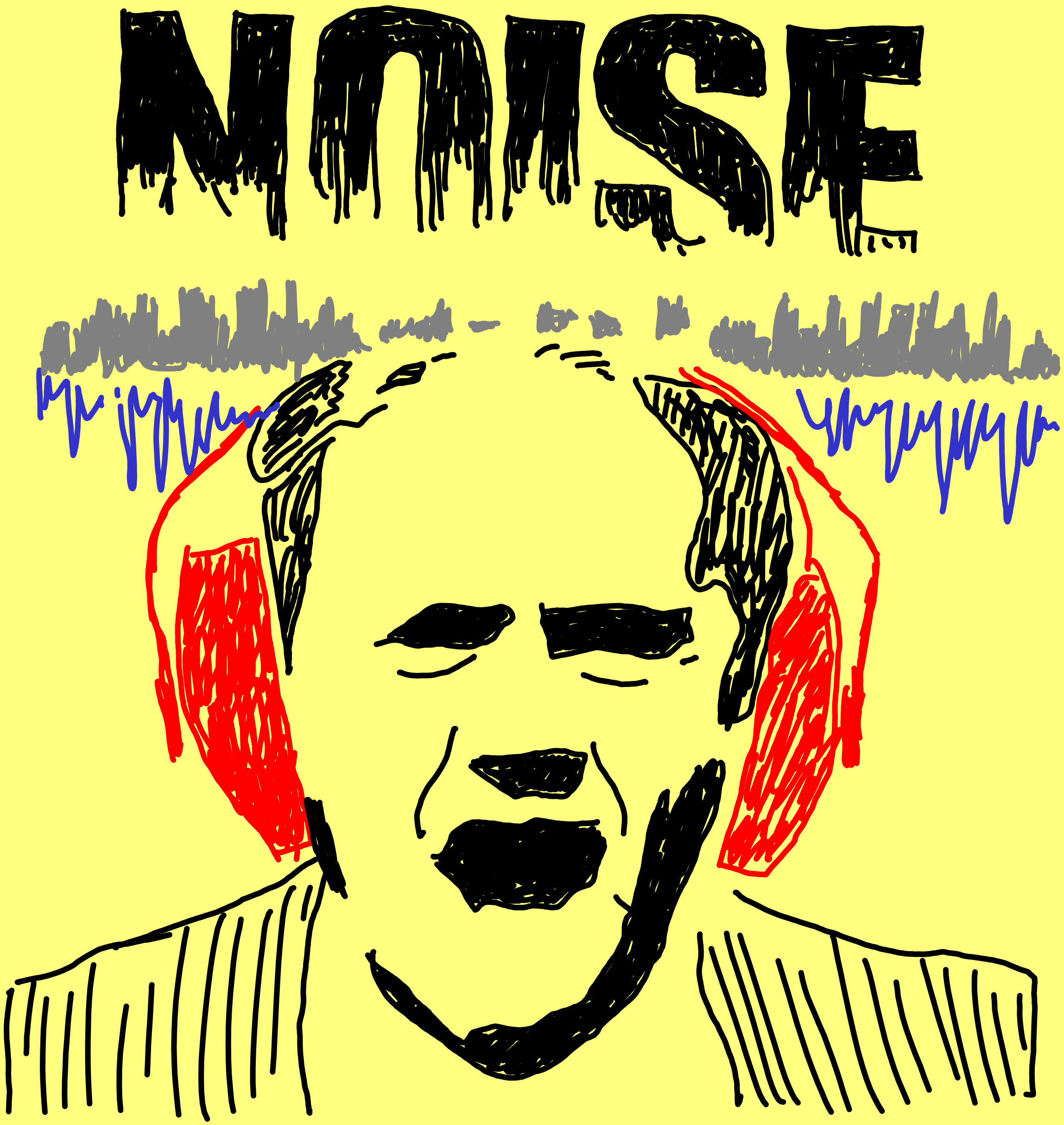
Measuring Scale before Simplification

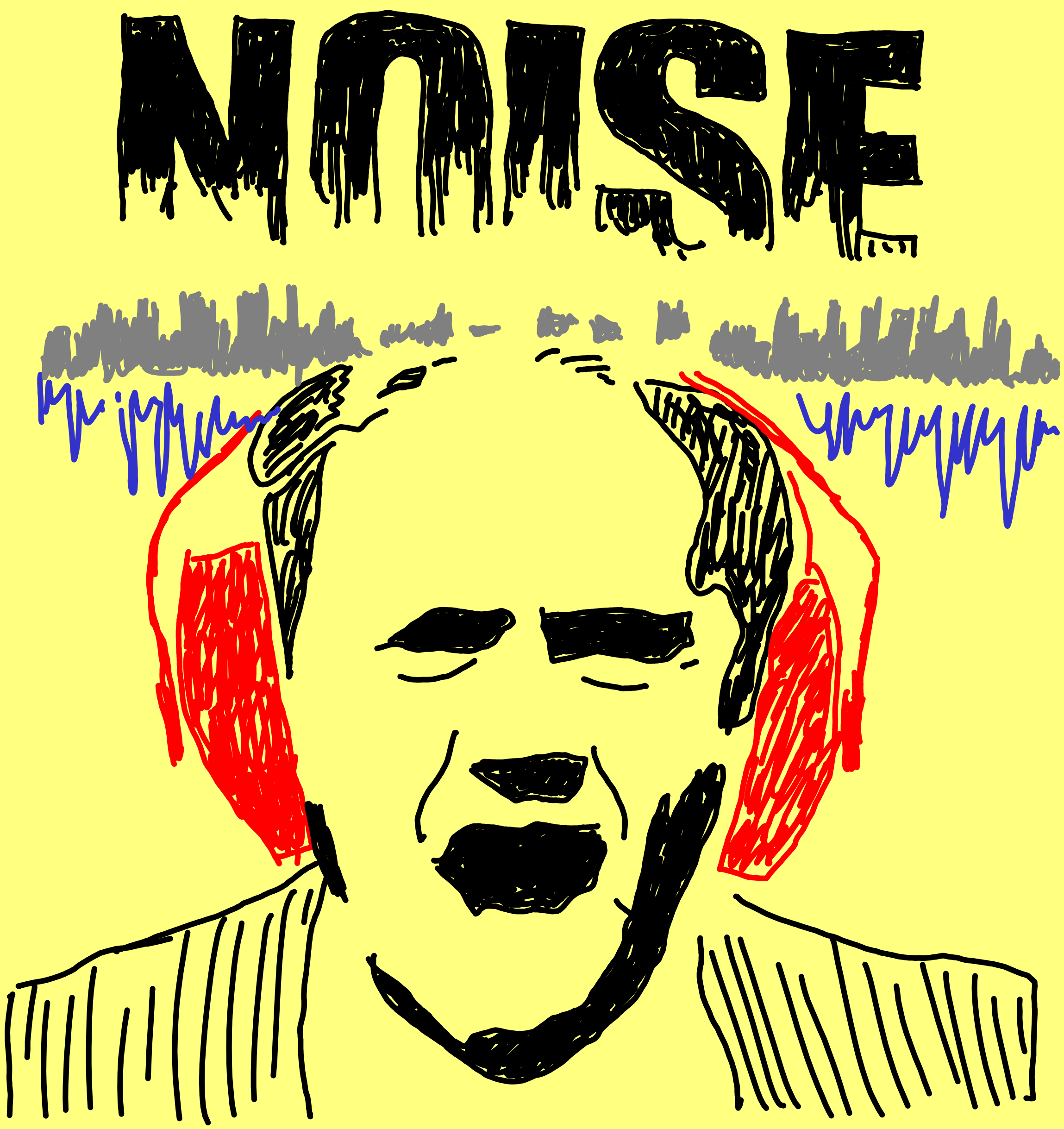
Columbia

Measuring Scale before Simplification

Herbert Edelsbrunner

Duke University
& Geomagic


NOISE



. . . IS ALWAYS AND EVERYWHERE.

Measure

Measure

. . . and don't simplify unless you must.

I Persistence

II Curves

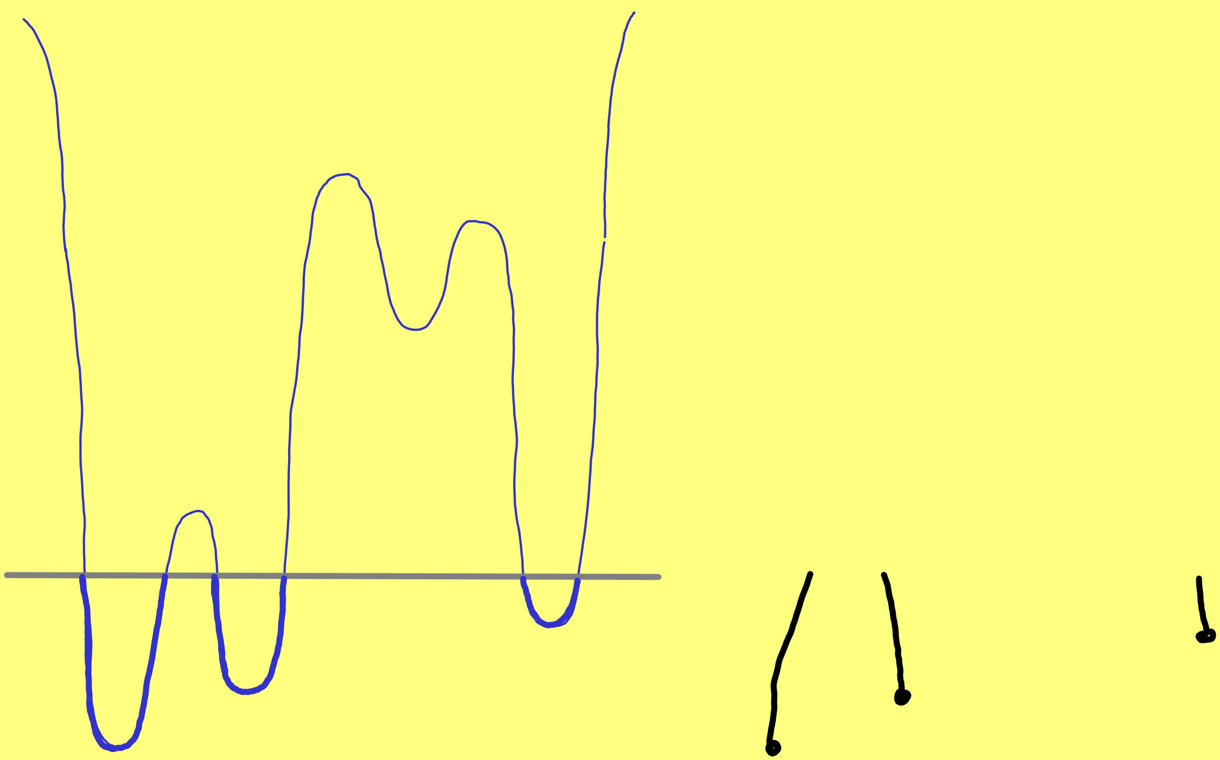
III Somites

I.1 One-D Functions

function

$f : \mathbb{S}^1 \to \mathbb{R}$

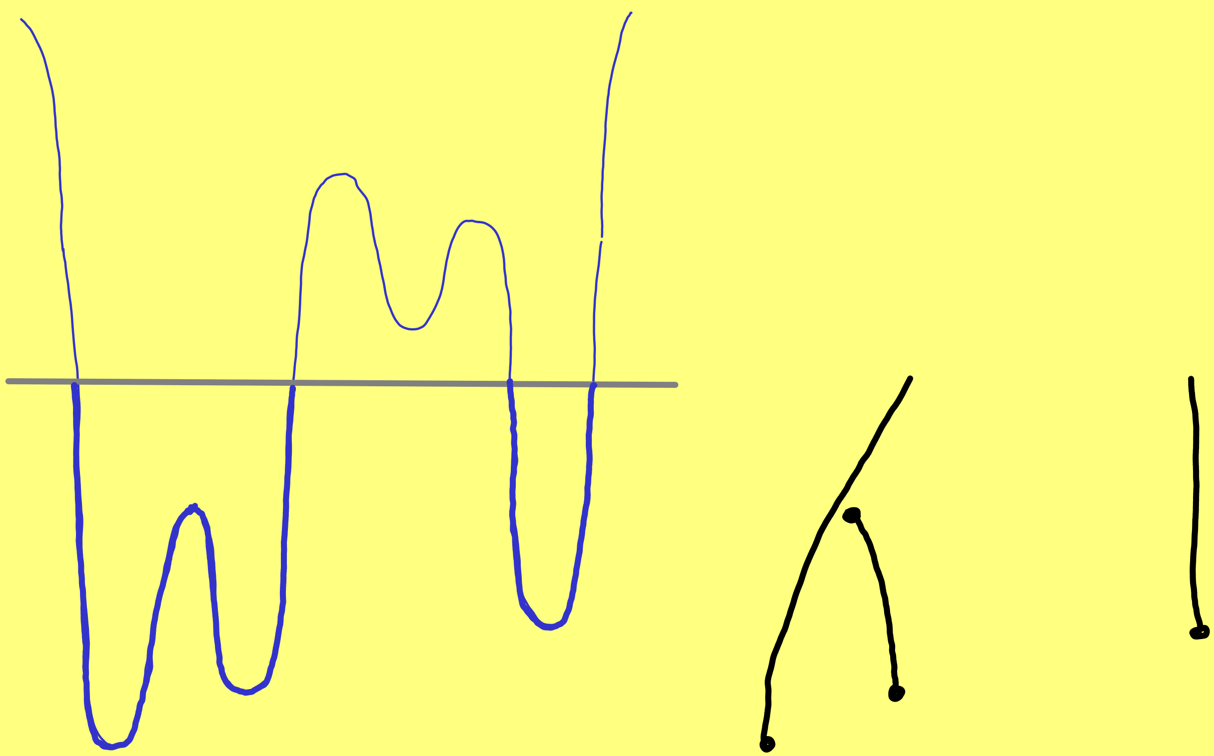
I.1 One-D Functions



function

$f: \mathbb{S}^1 \to \mathbb{R}$

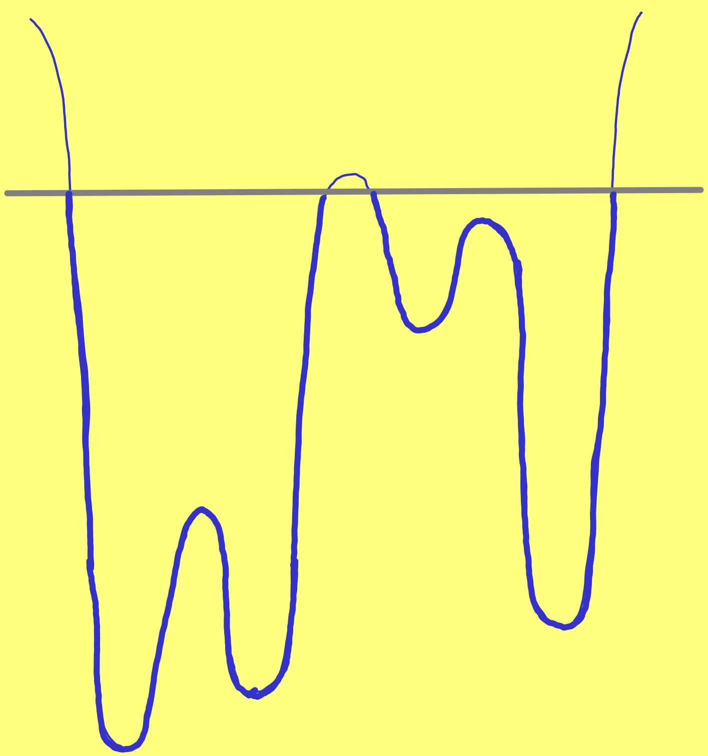
I.1 One-D Functions

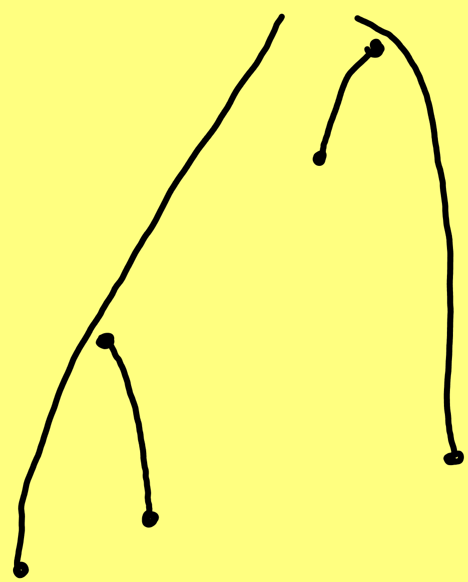


function

$f : S^1 \to \mathbb{R}$

I.1 ONE-D FUNCTIONS





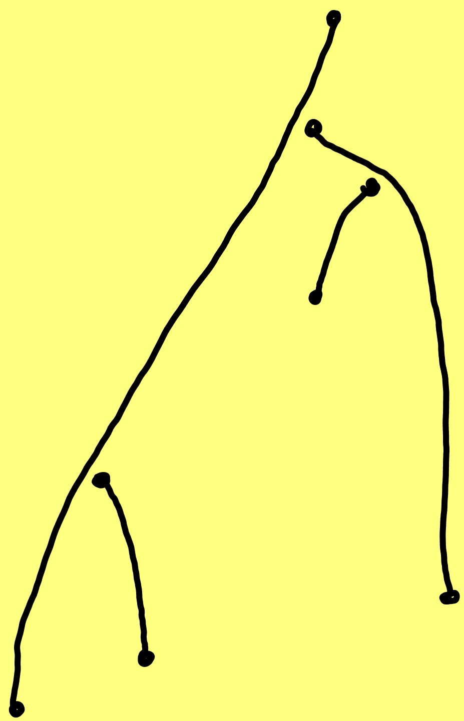
function

$f : \mathbb{S}^1 \to \mathbb{R}$

I.1 ONE-D FUNCTIONS

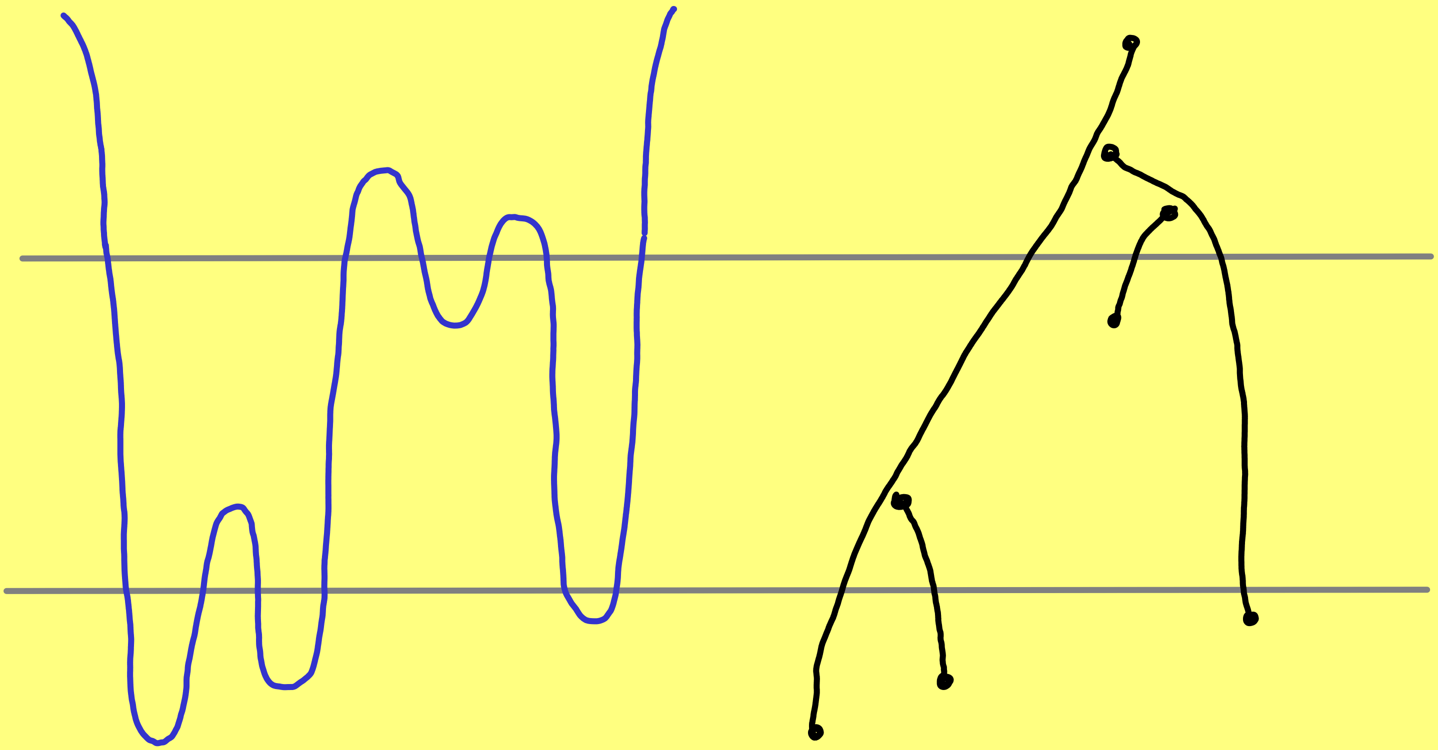
function

$f : \mathbb{S}^1 \to \mathbb{R}$



merge tree

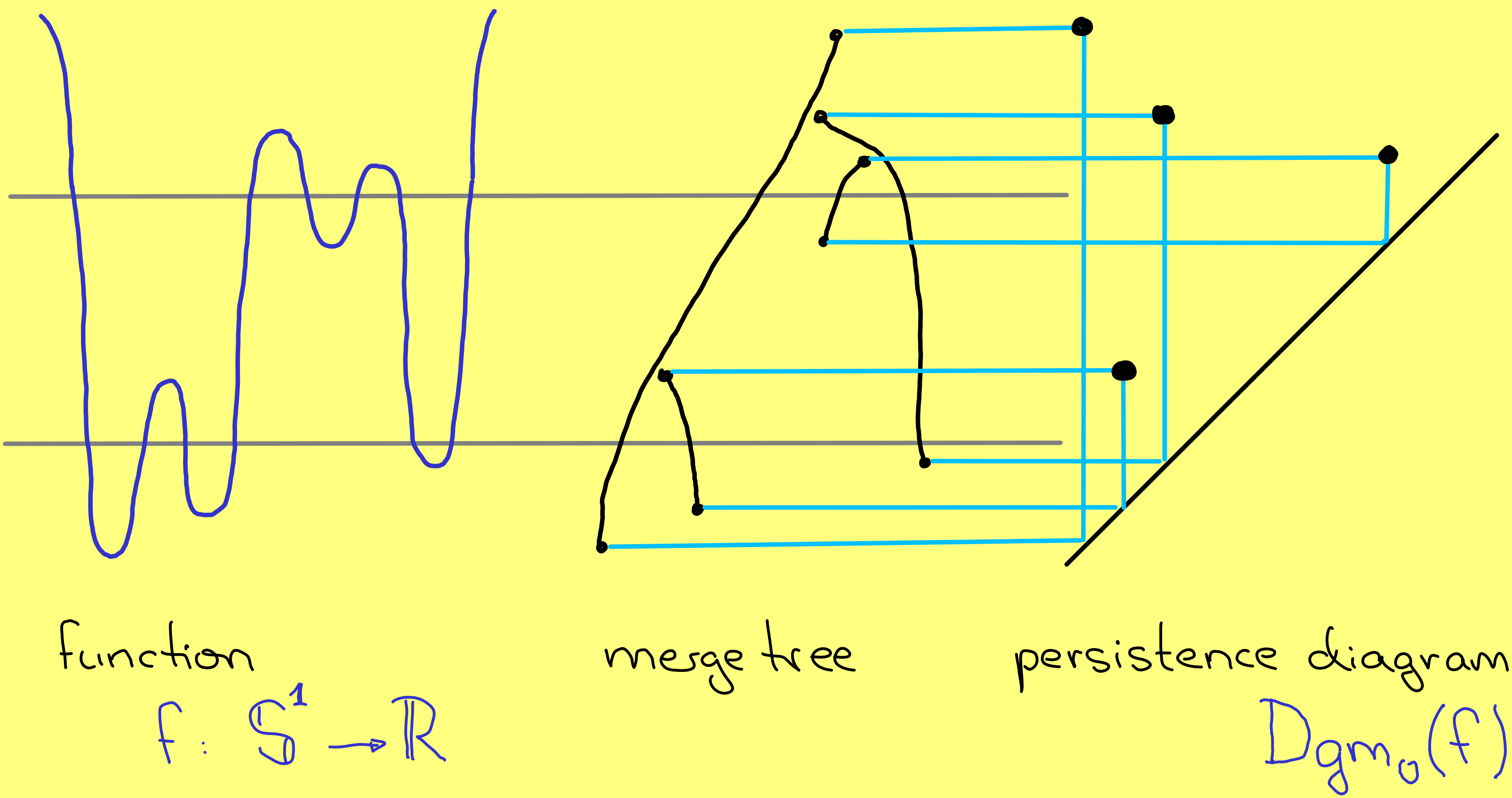
I.1 One-D Functions



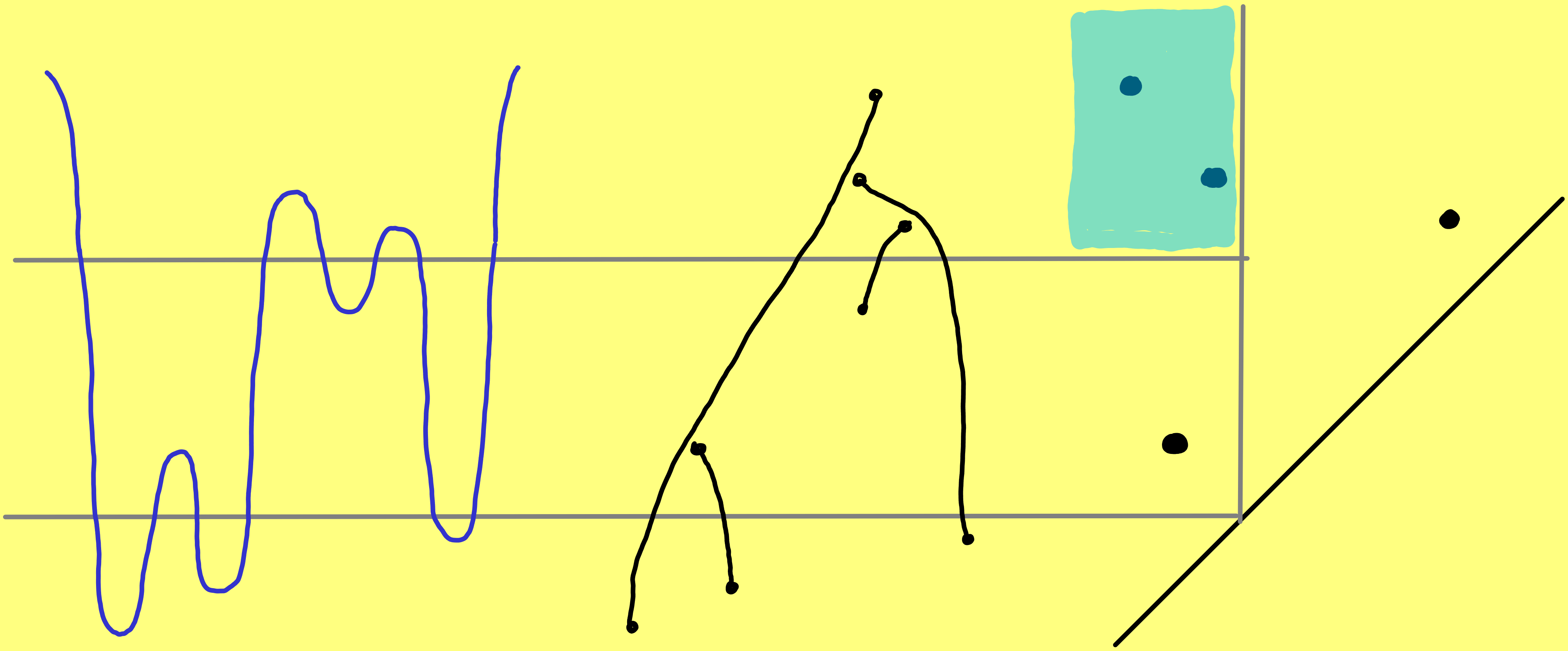
function $\quad\quad\quad\quad$ merge tree

$f : \mathbb{S}^1 \to \mathbb{R}$

I.1 One-D Functions



I.1 One-D Functions



function

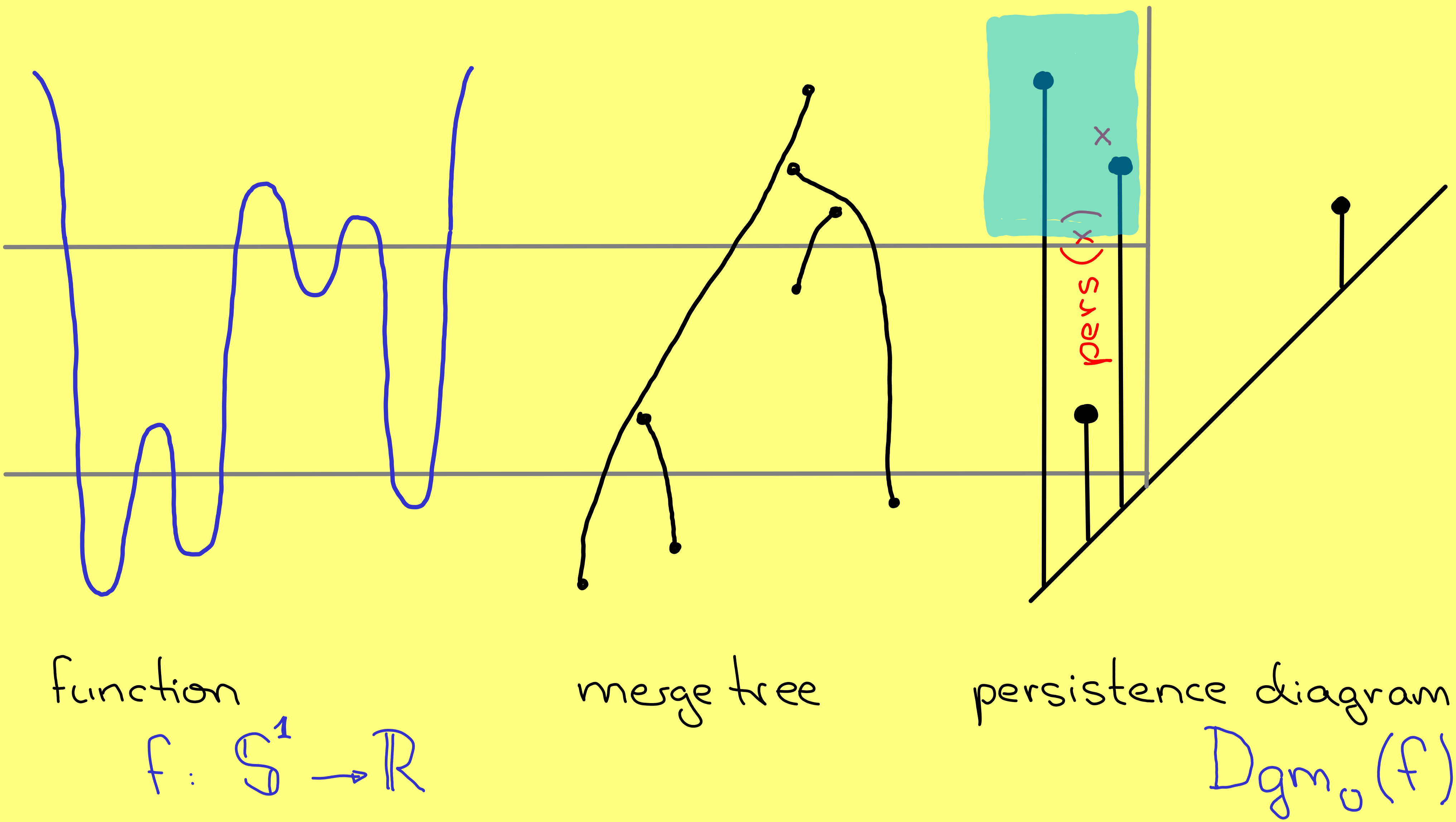
$f : \mathbb{S}^1 \to \mathbb{R}$

merge tree

persistence diagram

$\mathrm{Dgm}_0(f)$

I.1 One-D Functions

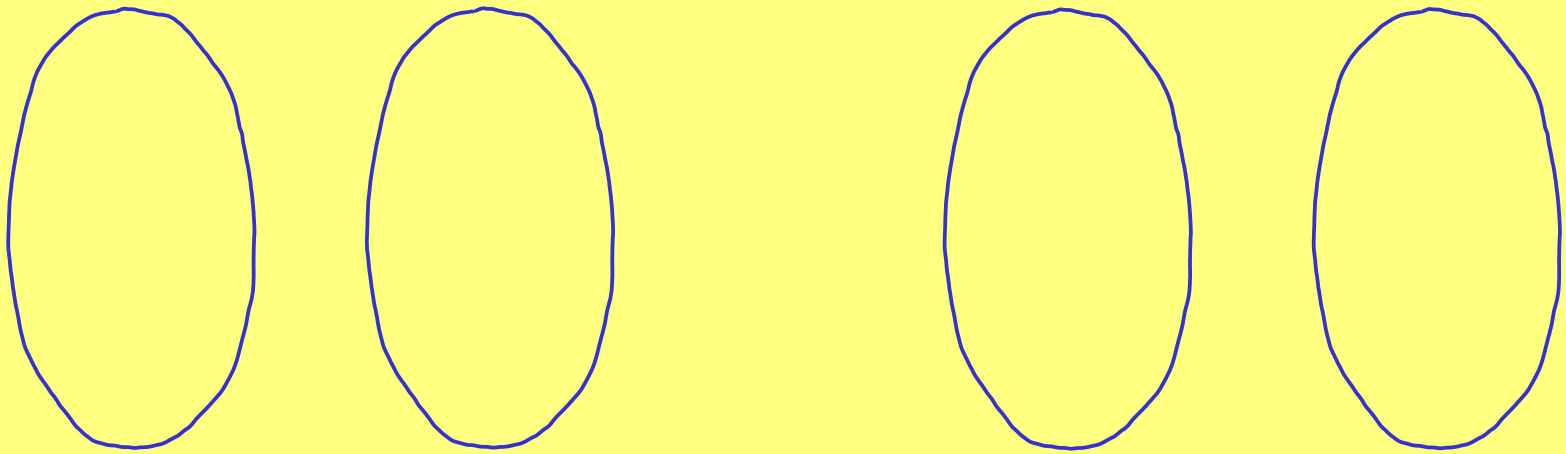


I.2 Filtrations

$$\ldots \subseteq K_{i-1} \subseteq K_i \subseteq \ldots \subseteq K_{j-1} \subseteq K_j \subseteq \ldots$$

I.2 Filtrations

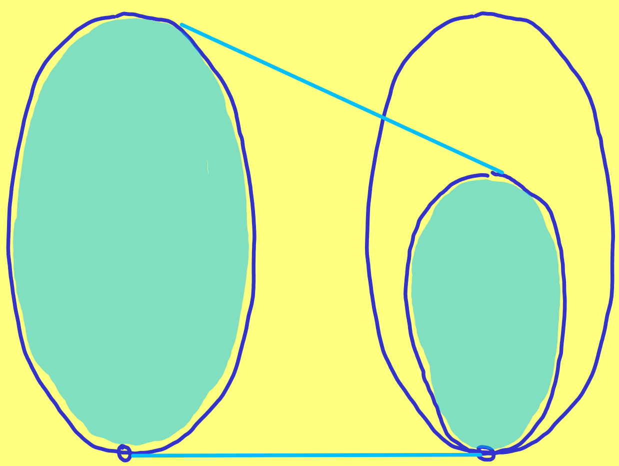
$$\cdots \subseteq K_{i-1} \subseteq K_i \subseteq \cdots \subseteq K_{j-1} \subseteq K_j \subseteq \cdots$$



$$\cdots \longrightarrow H_p(K_{i-1}) \longrightarrow H_p(K_i) \longrightarrow \cdots \longrightarrow H_p(K_{j-1}) \longrightarrow H_p(K_j) \longrightarrow \cdots$$

I.2 Filtrations

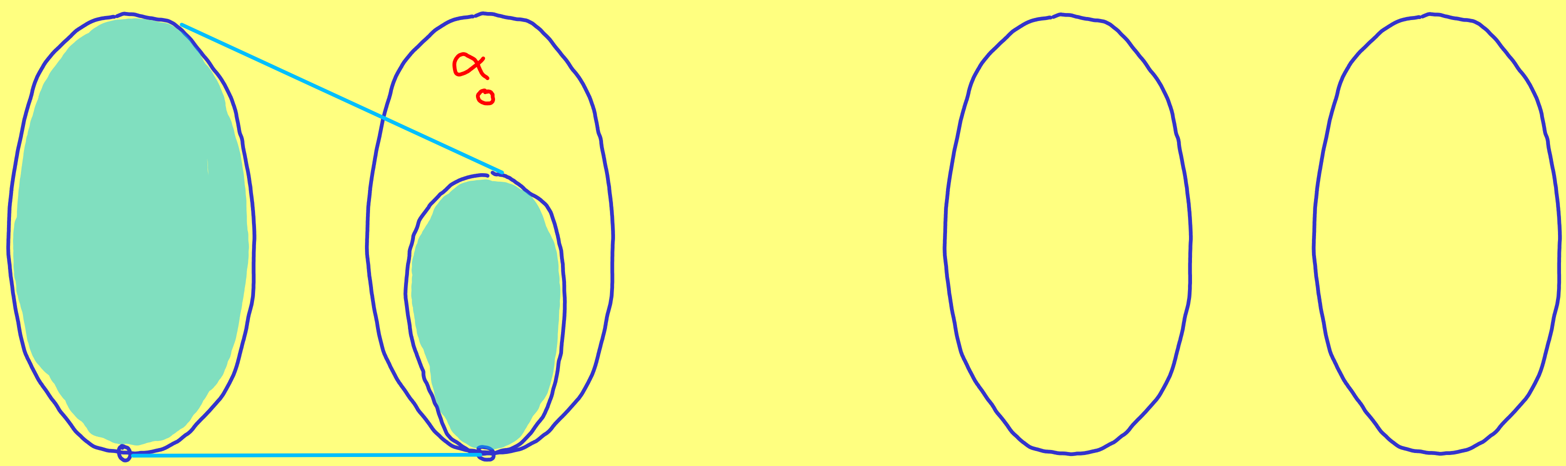
$$\cdots \subseteq K_{i-1} \subseteq K_i \subseteq \cdots \subseteq K_{j-1} \subseteq K_j \subseteq \cdots$$



$$\cdots \longrightarrow H_p(K_{i-1}) \longrightarrow H_p(K_i) \longrightarrow \cdots \longrightarrow H_p(K_{j-1}) \longrightarrow H_p(K_j) \longrightarrow \cdots$$

I.2 Filtrations

$$\ldots \subseteq K_{i-1} \subseteq K_i \subseteq \ldots \subseteq K_{j-1} \subseteq K_j \subseteq \ldots$$

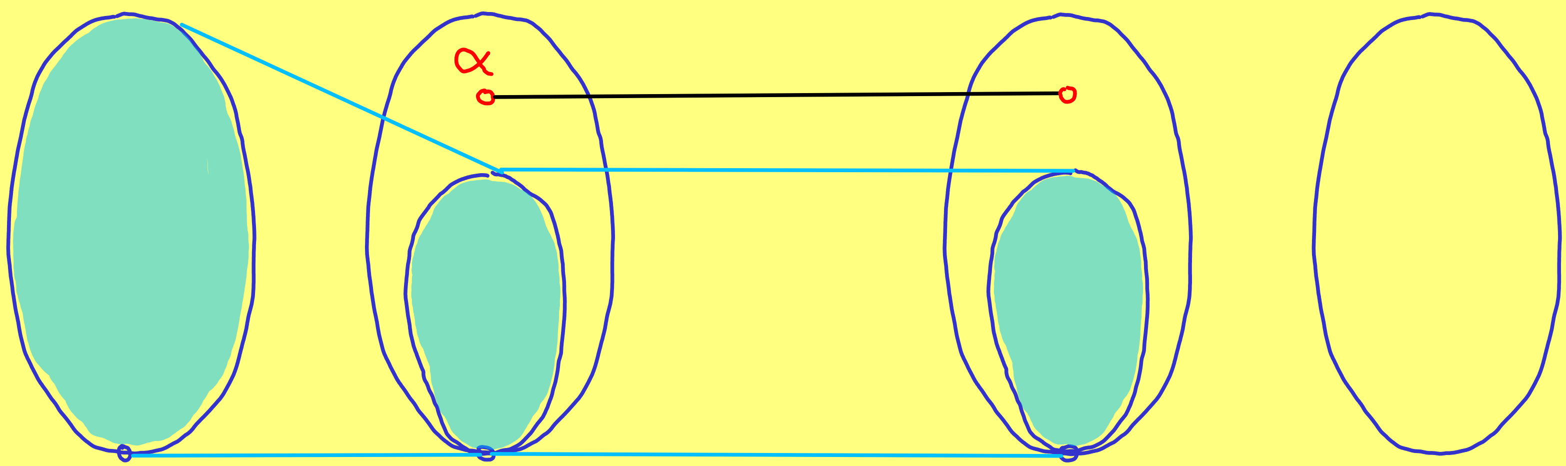


$$\ldots \longrightarrow H_p(K_{i-1}) \longrightarrow H_p(K_i) \longrightarrow \ldots \longrightarrow H_p(K_{j-1}) \longrightarrow H_p(K_j) \longrightarrow \ldots$$

α is born at K_i

I.2 FILTRATIONS

$$\dots \subseteq K_{i-1} \subseteq K_i \subseteq \dots \subseteq K_{j-1} \subseteq K_j \subseteq \dots$$

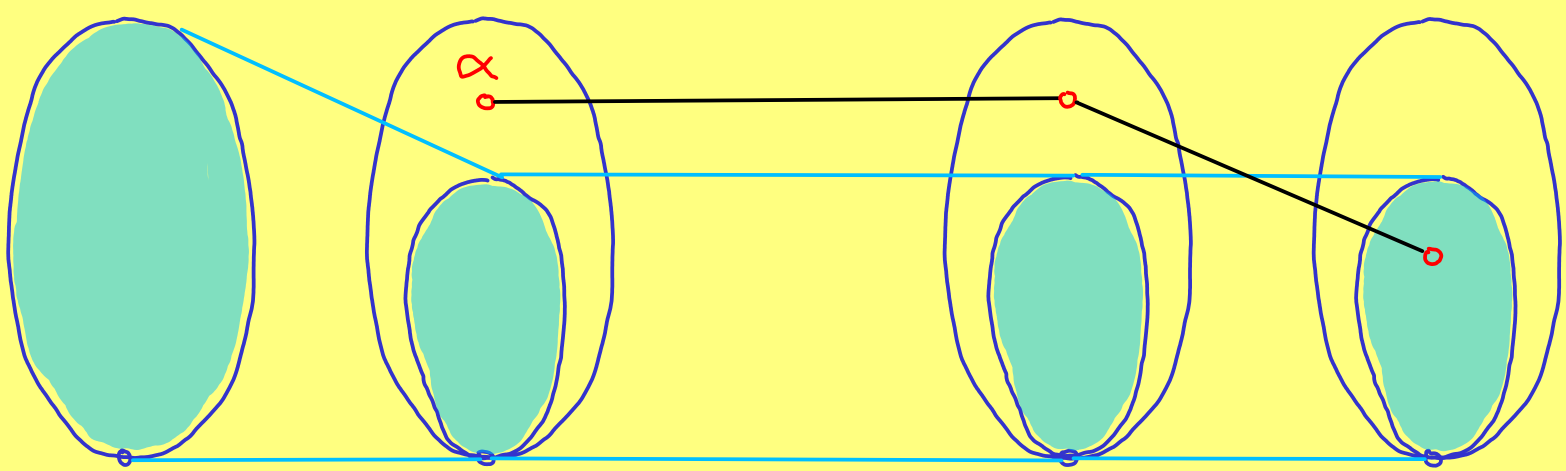


$$\dots \longrightarrow H_p(K_{i-1}) \longrightarrow H_p(K_i) \longrightarrow \dots \longrightarrow H_p(K_{j-1}) \longrightarrow H_p(K_j) \longrightarrow \dots$$

α is born at K_i

I.2 Filtrations

$$\ldots \subseteq K_{i-1} \subseteq K_i \subseteq \ldots \subseteq K_{j-1} \subseteq K_j \subseteq \ldots$$

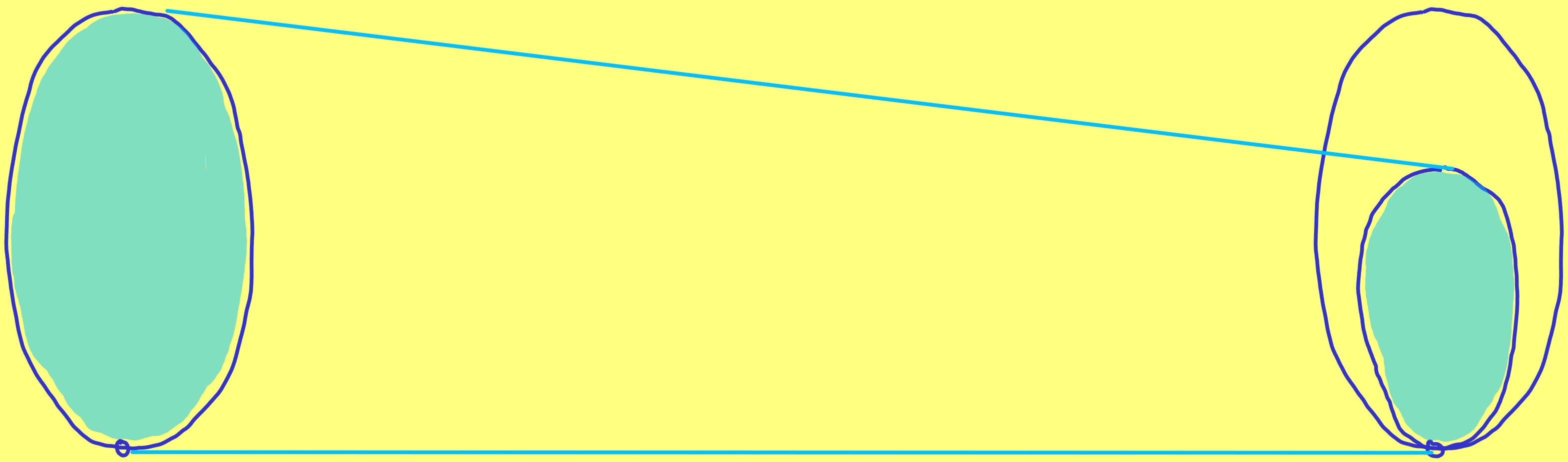


$$\ldots \to H_p(K_{i-1}) \to H_p(K_i) \to \ldots \to H_p(K_{j-1}) \to H_p(K_j) \to \ldots$$

α is born at K_i and dies entering K_j

I.2 Filtrations

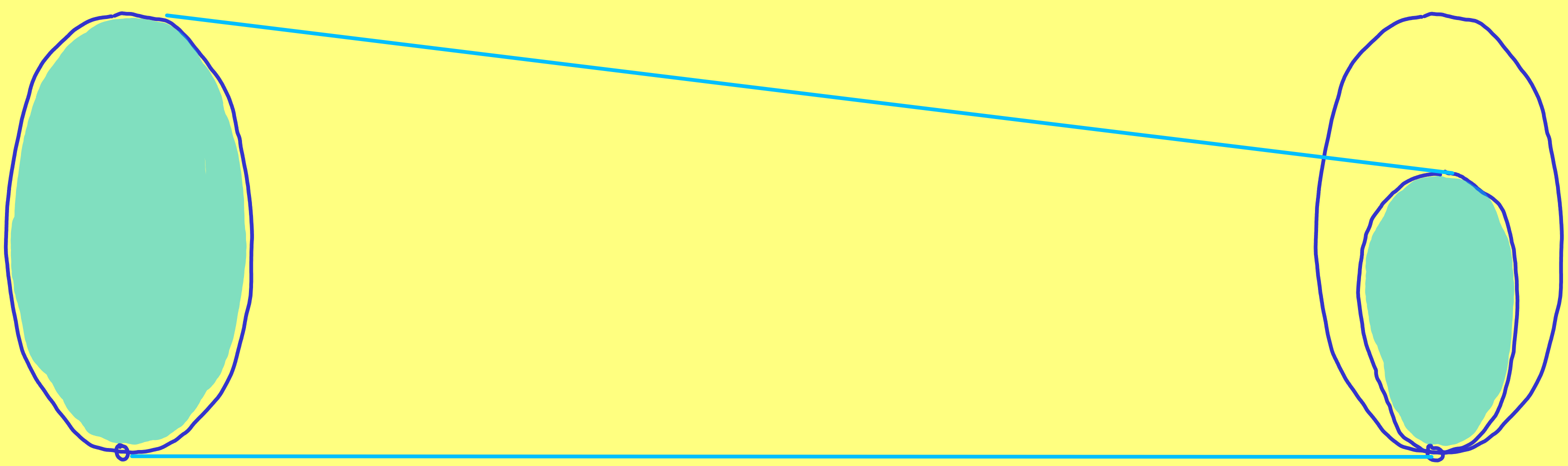
$$\dots \subseteq K_{i-1} \subseteq K_i \subseteq \dots \subseteq K_{j-1} \subseteq K_j \subseteq \dots$$



$$H_p(K_{i-1}) \to \quad \dots \quad \to H_p(K_j)$$

I.2 FILTRATIONS

$$\ldots \subseteq K_{i-1} \subseteq K_i \subseteq \ldots \subseteq K_{j-1} \subseteq K_j \subseteq \ldots$$

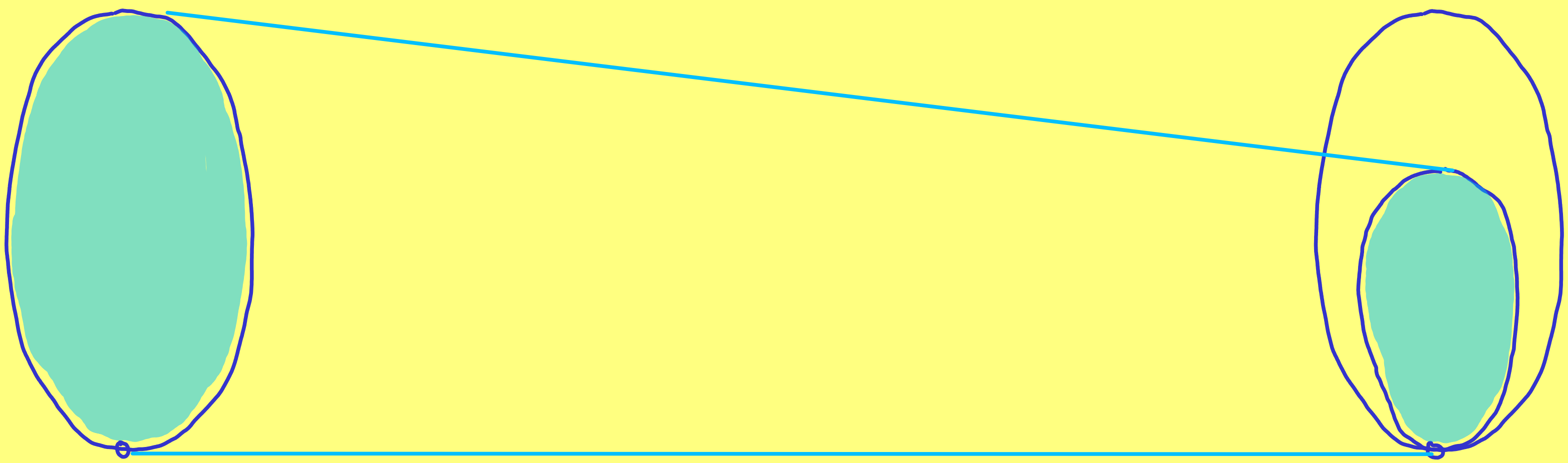


$$h_p^{i-1,j} : H_p(K_{i-1}) \longrightarrow \quad \ldots \quad \longrightarrow H_p(K_j)$$

induced homomorphism

I.2 Filtrations

$$\ldots \subseteq K_{i-1} \subseteq K_i \subseteq \ldots \subseteq K_{j-1} \subseteq K_j \subseteq \ldots$$



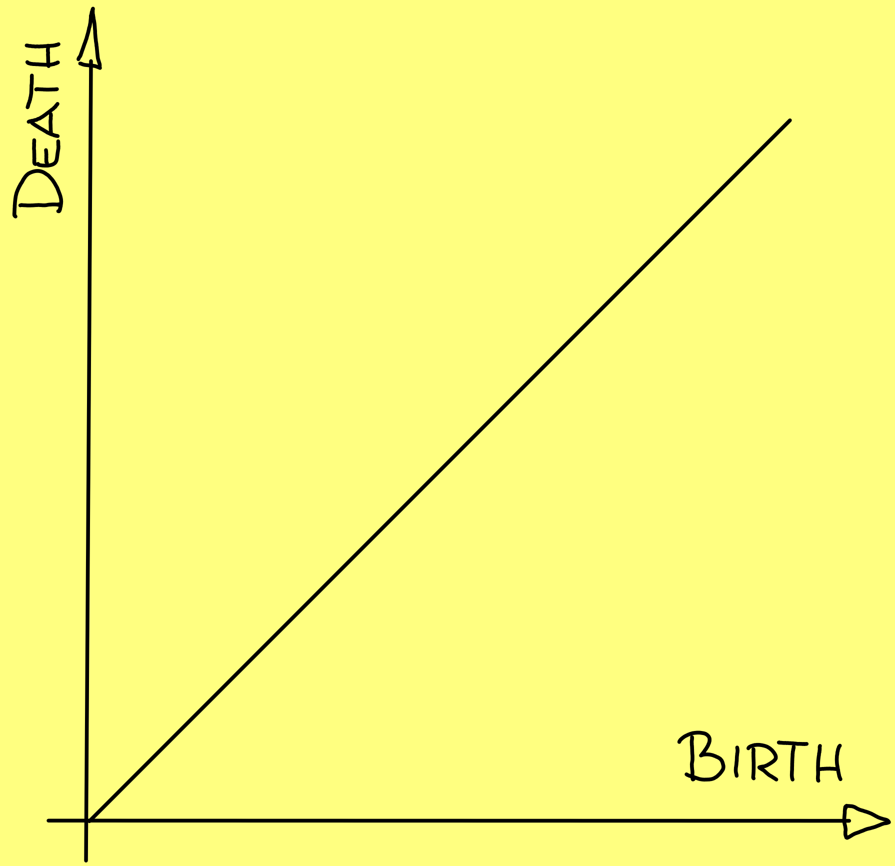
$$h_p^{i-1,j} : H_p(K_{i-1}) \to \ldots \to H_p(K_j)$$

induced homomorphism

$\mathrm{im}\, h_p^{i-1,j}$ is persistent homology group

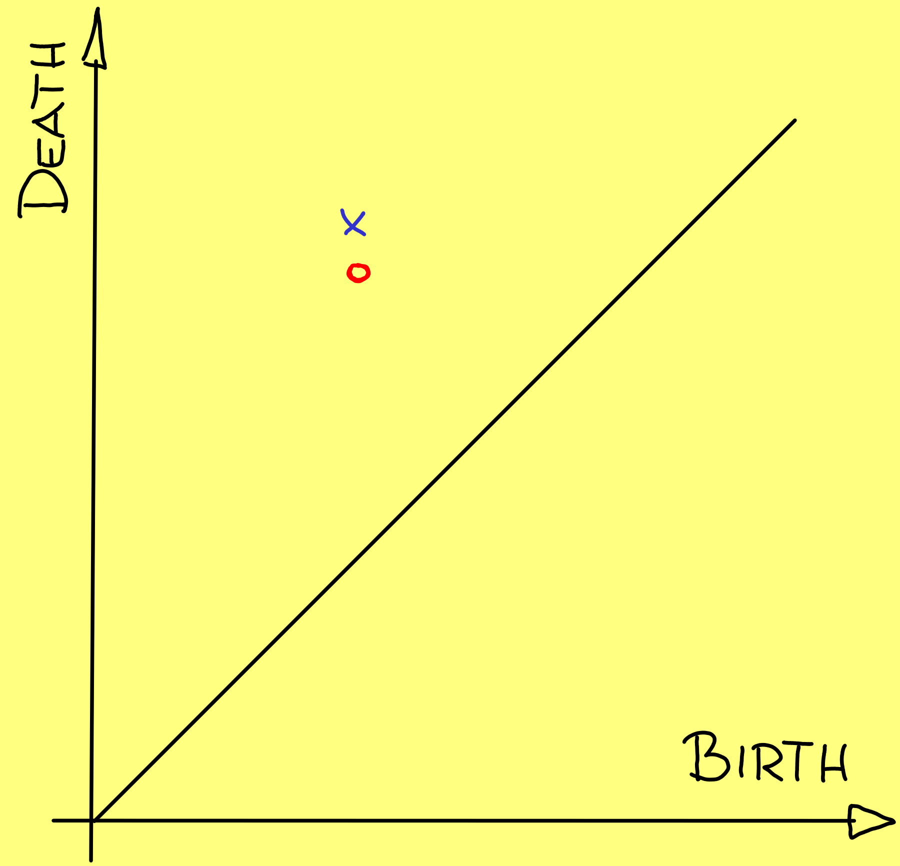
I.3 Diagrams

$\mathrm{Dgm}_p(f)$:



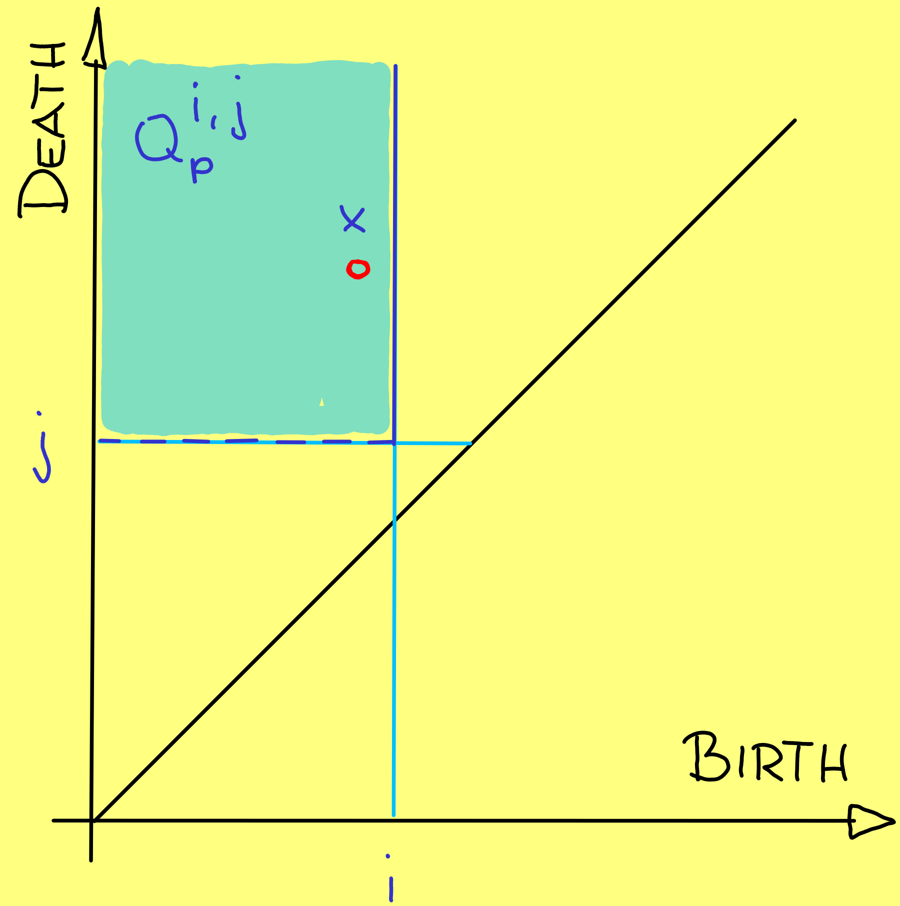
I.3 Diagrams

$\mathrm{Dgm}_p(f)$:



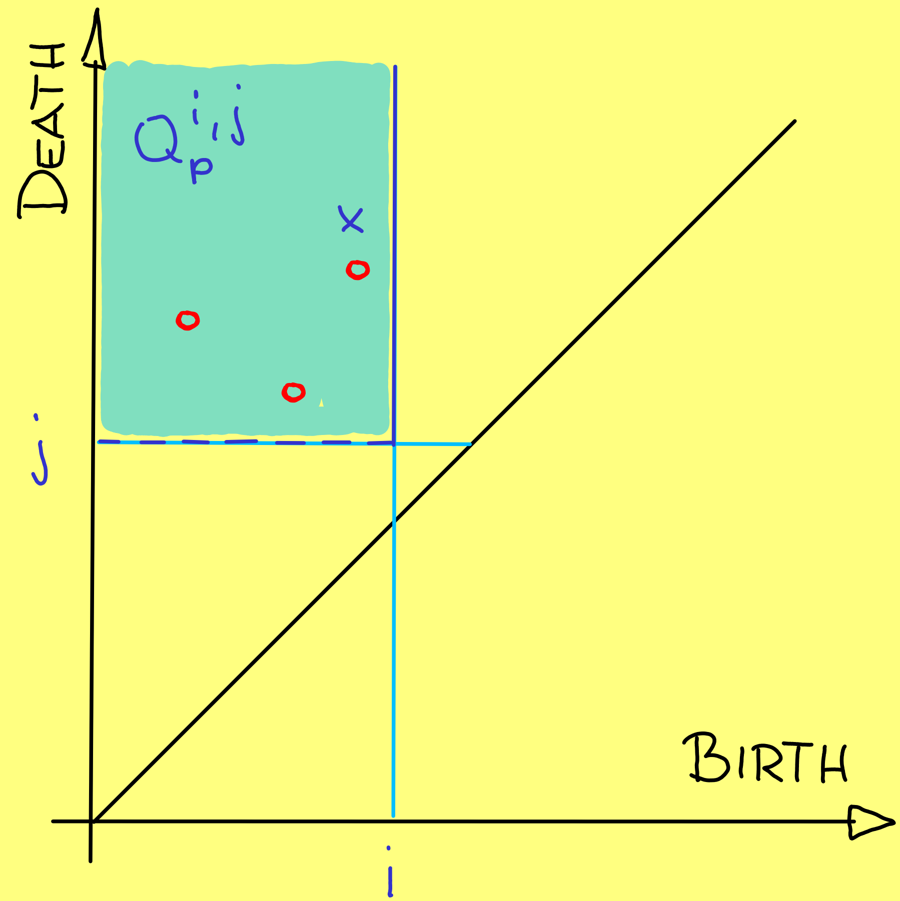
I.3 DIAGRAMS

$Dgm_p(f)$:



I.3 DIAGRAMS

$Dgm_p(f)$:



FUNDAMENTAL LEMMA OF PERSISTENT HOMOLOGY.

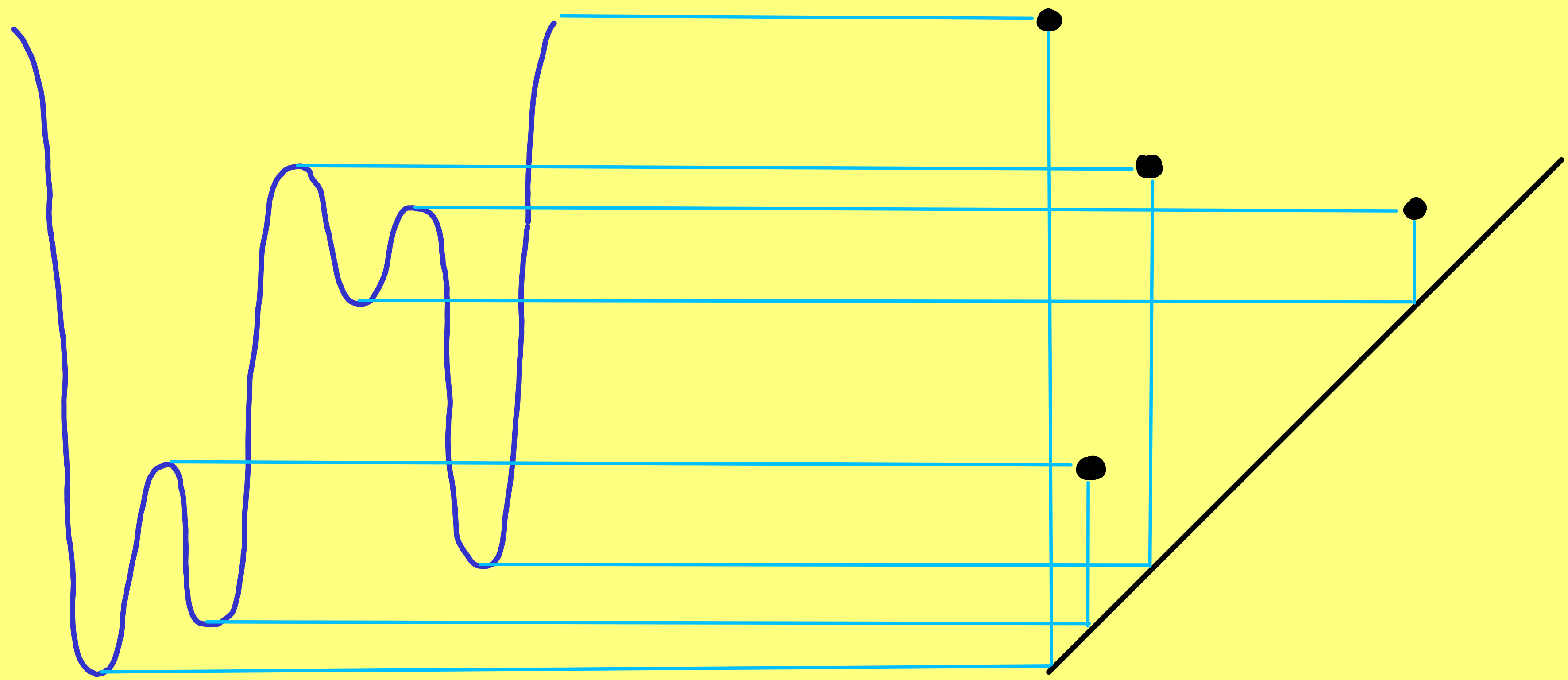
#points in $Q_p^{i,j}$ = rank im $h_p^{i,j}$.

I Persistence

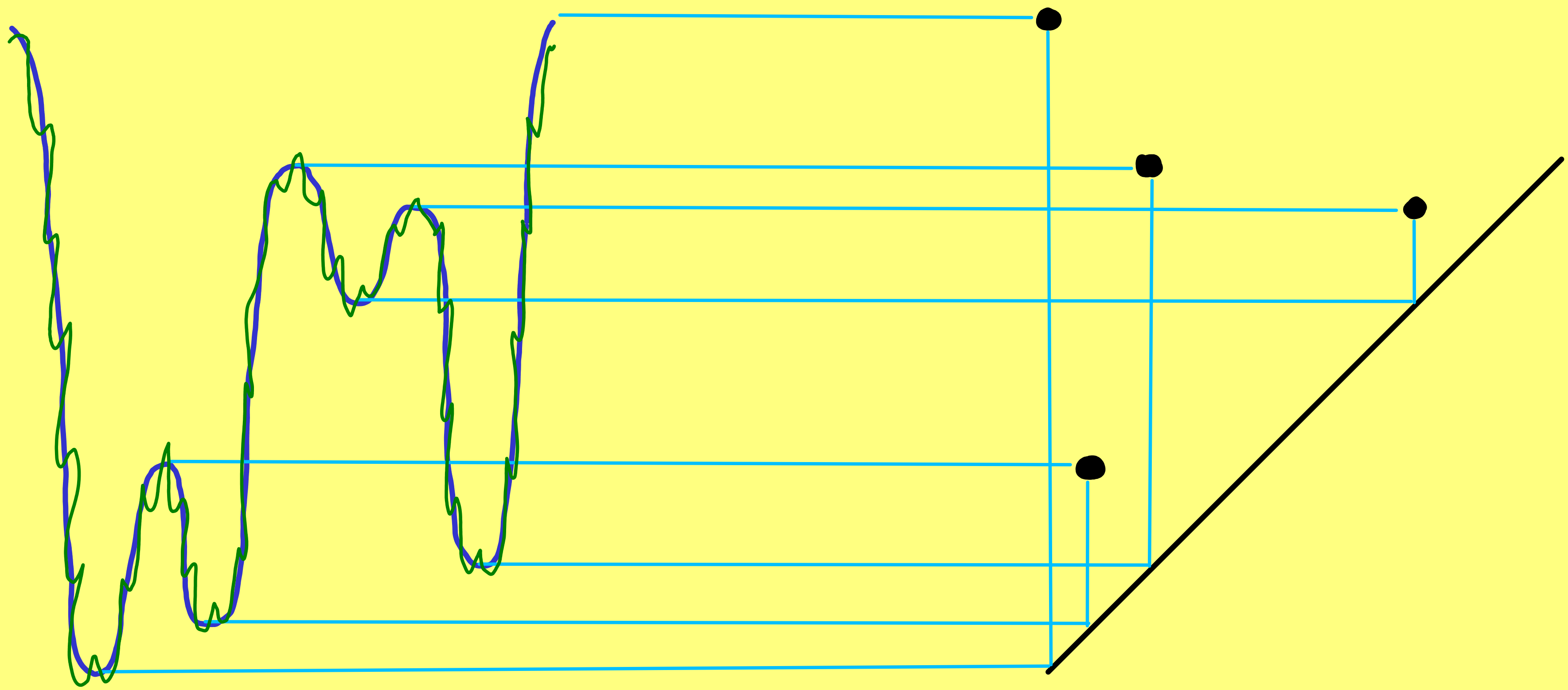
II Curves

III Somites

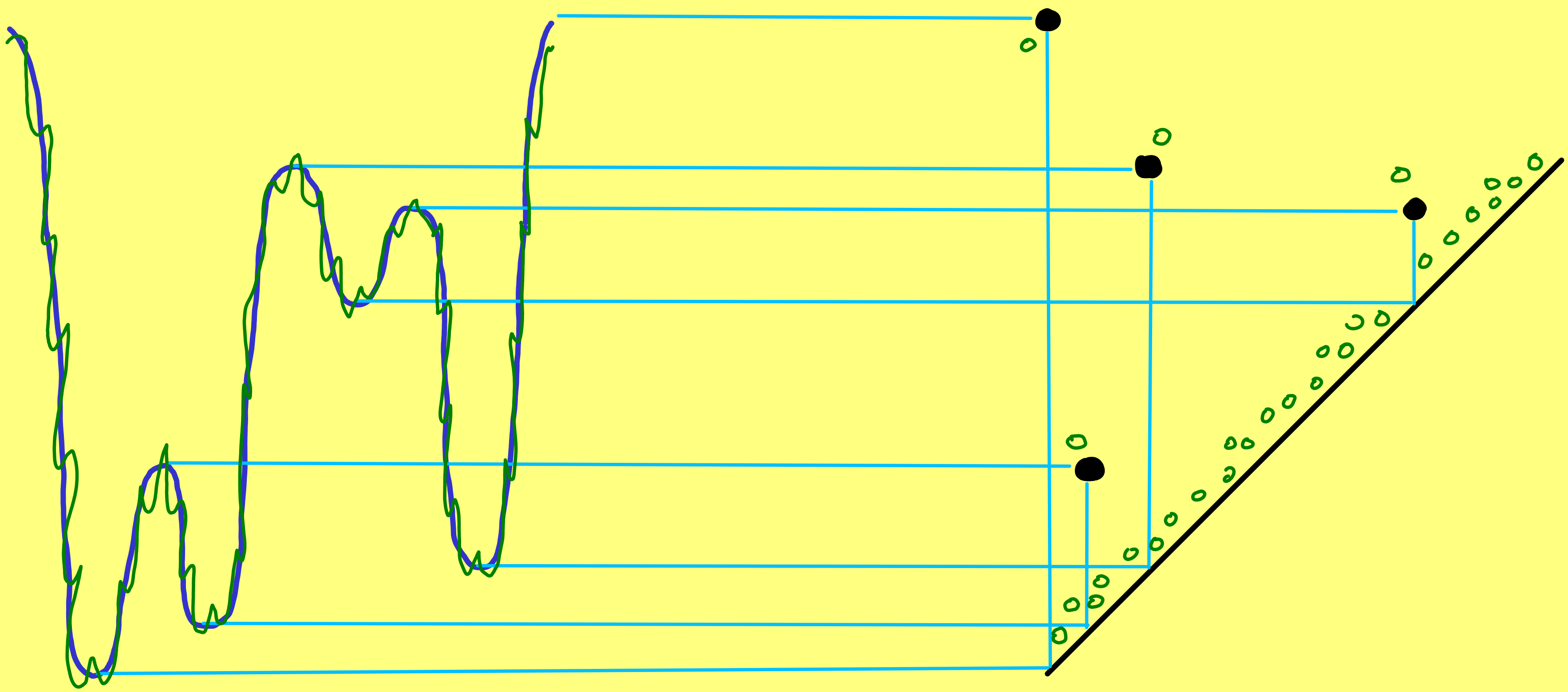
II.1 Bottleneck Distance



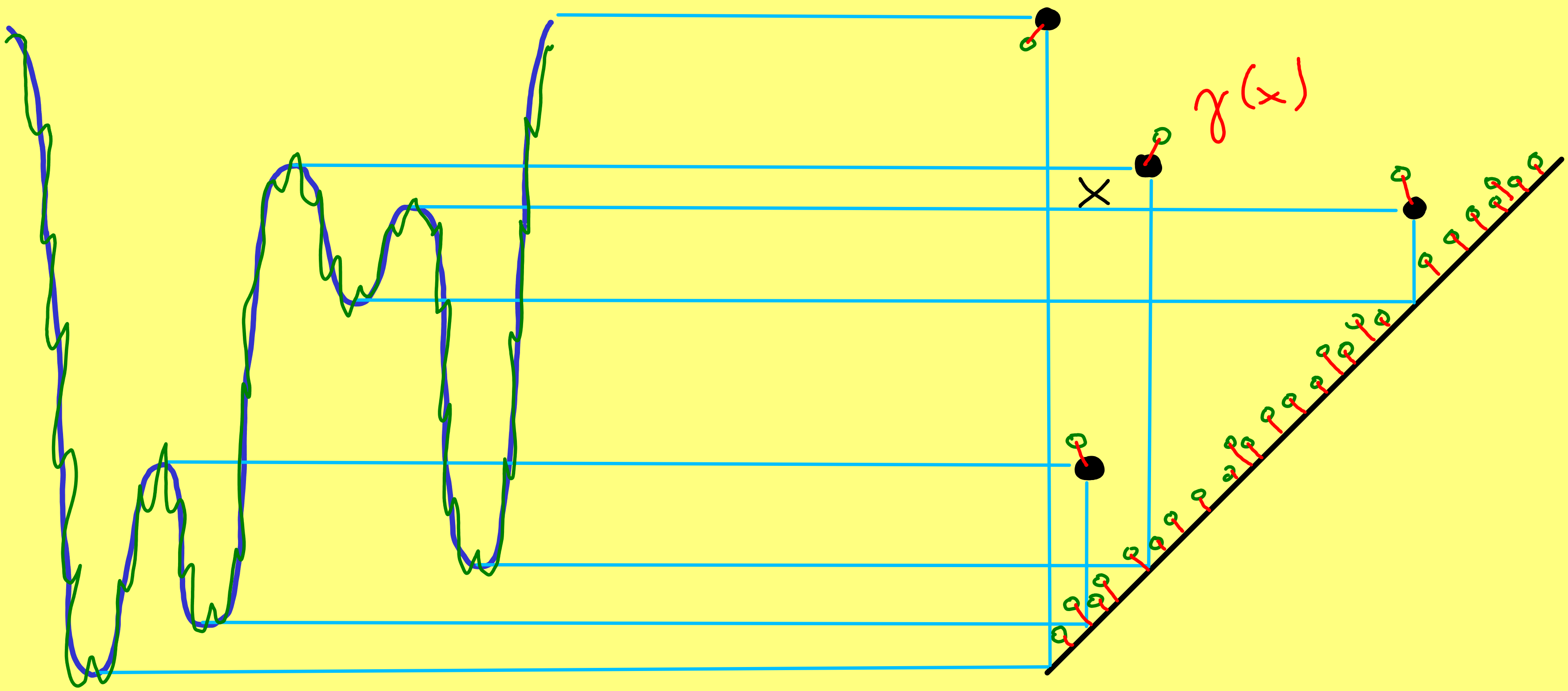
II.1 Bottleneck Distance



II.1 Bottleneck Distance



II.1 Bottleneck Distance



The bottleneck distance between the diagrams is

$$W_\infty\left(\mathrm{Dgm}_p(f), \mathrm{Dgm}_p(g)\right) = \inf_{\gamma} \sup_{x} \|x - \gamma(x)\|_\infty .$$

II.1 BOTTLENECK DISTANCE

L_∞-STAB. THM. For tame functions $f, g : \mathbb{X} \to \mathbb{R}$ the bottleneck distance between their diagrams is bounded by the max-difference between the functions,

$$W_\infty\left(\mathrm{Dgm}_p(f), \mathrm{Dgm}_p(g)\right) \leq \|f-g\|_\infty .$$

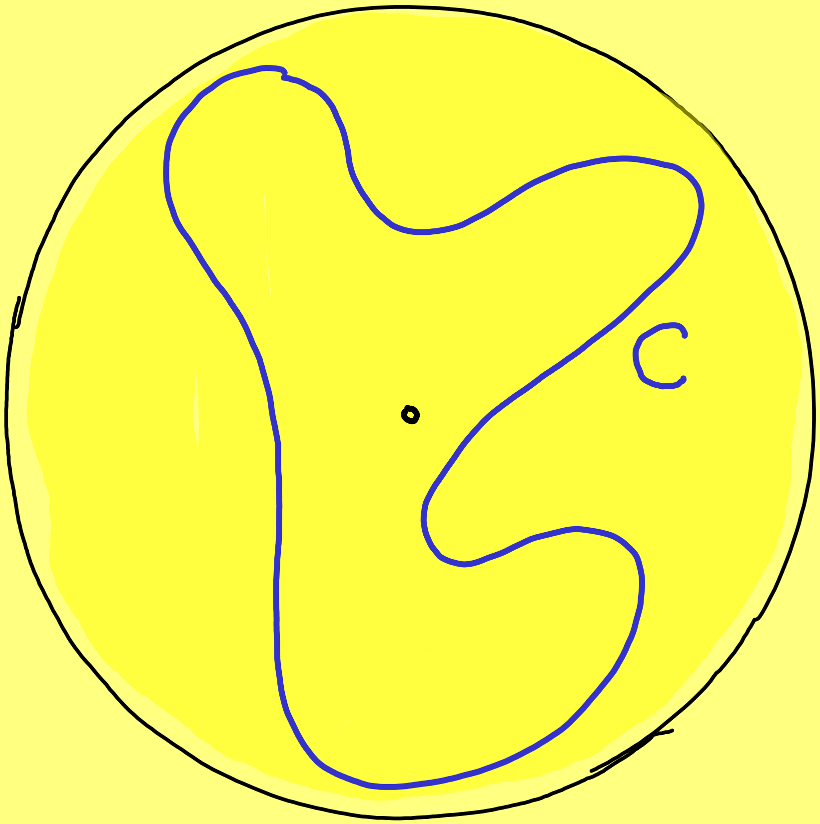
II.1 BOTTLENECK DISTANCE

L_∞-STAB. THM. For tame functions $f, g : \mathbb{X} \to \mathbb{R}$ the bottleneck distance between their diagrams is bounded by the max-difference between the functions,

$$W_\infty(\mathrm{Dgm}_p(f), \mathrm{Dgm}_p(g)) \leq \|f - g\|_\infty .$$

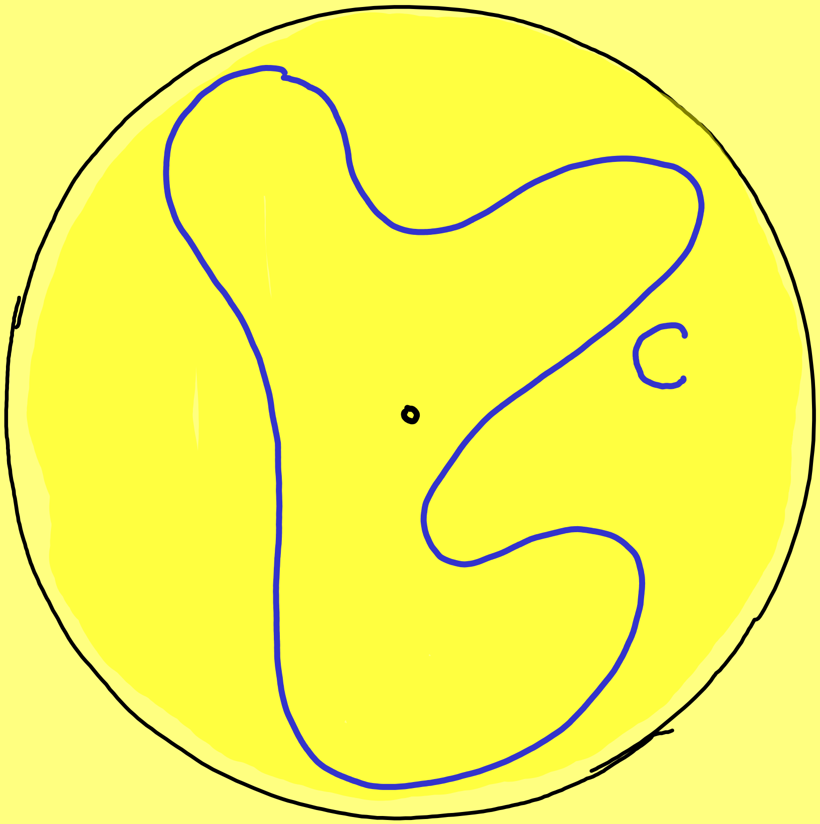
2004 Cohen-Steiner, E., Harer

II.2 FÁRY THEOREM



C = smooth curve inside unit disk

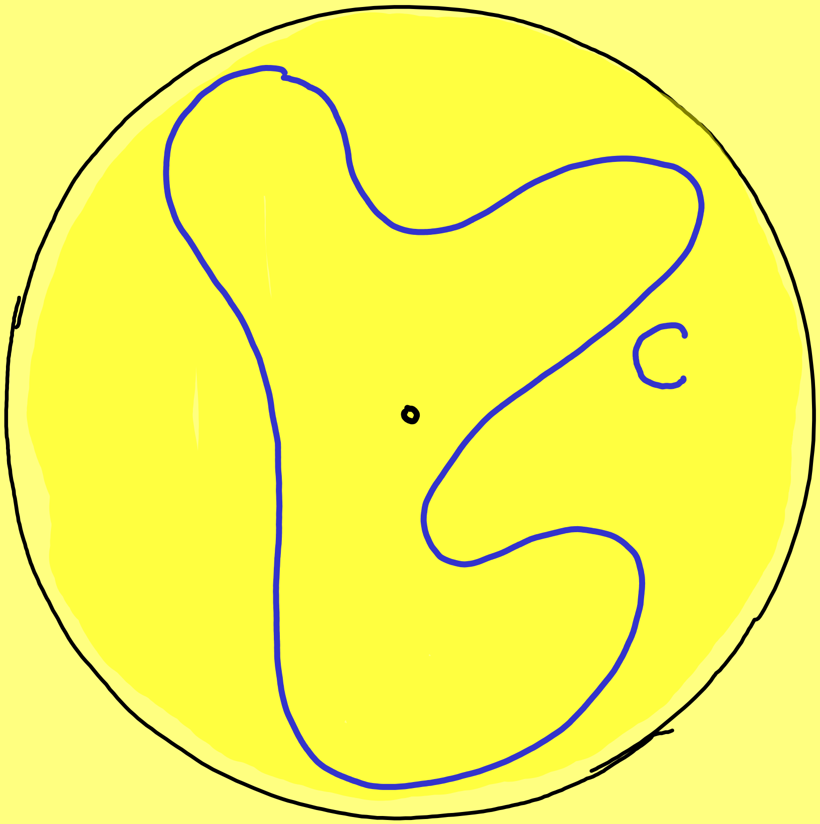
II.2 FÁRY THEOREM



C = smooth curve inside unit disk

$L(C)$ = length

II.2 FÁRY THEOREM

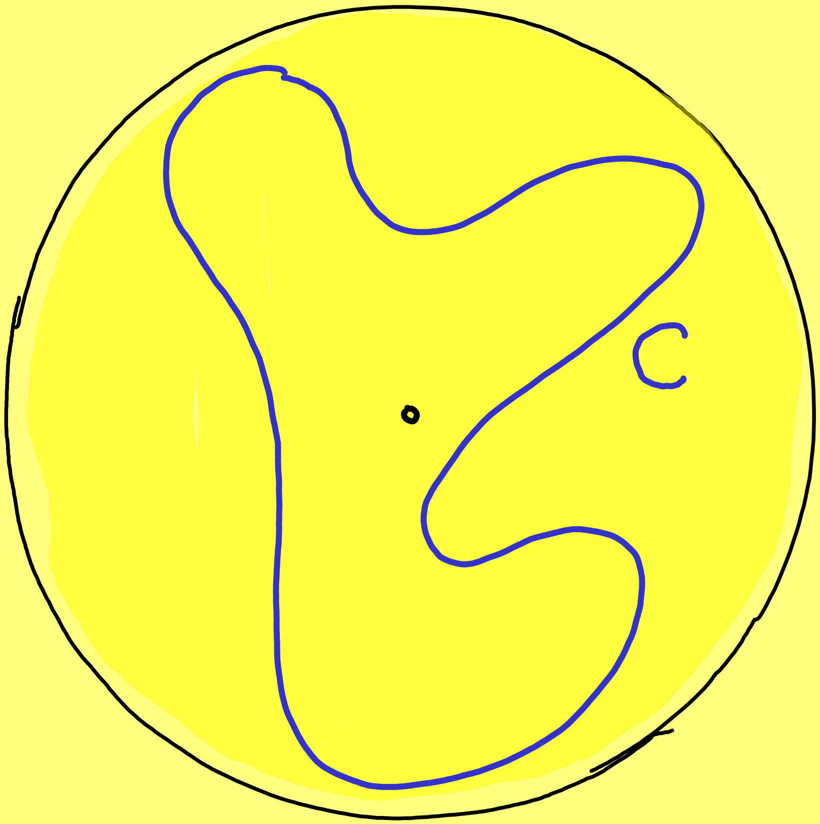


C = smooth curve inside unit disk

$L(C)$ = length

$K(C)$ = total curvature

II.2 FÁRY THEOREM



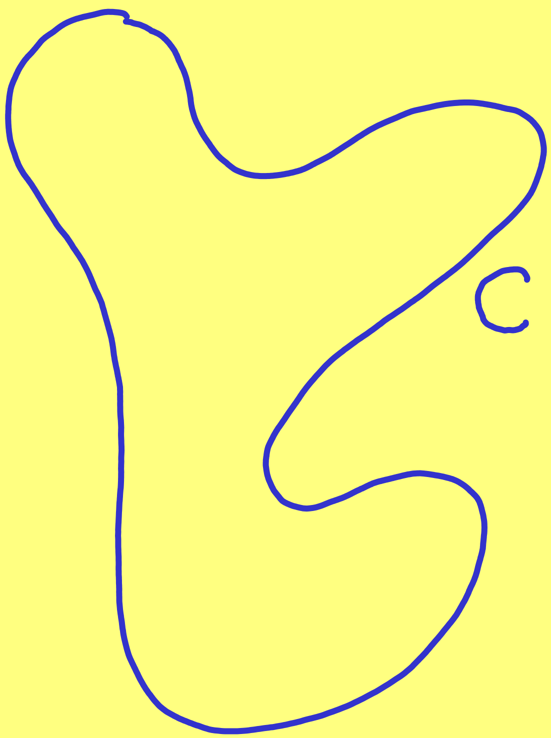
C = smooth curve inside unit disk

$L(C)$ = length

$K(C)$ = total curvature

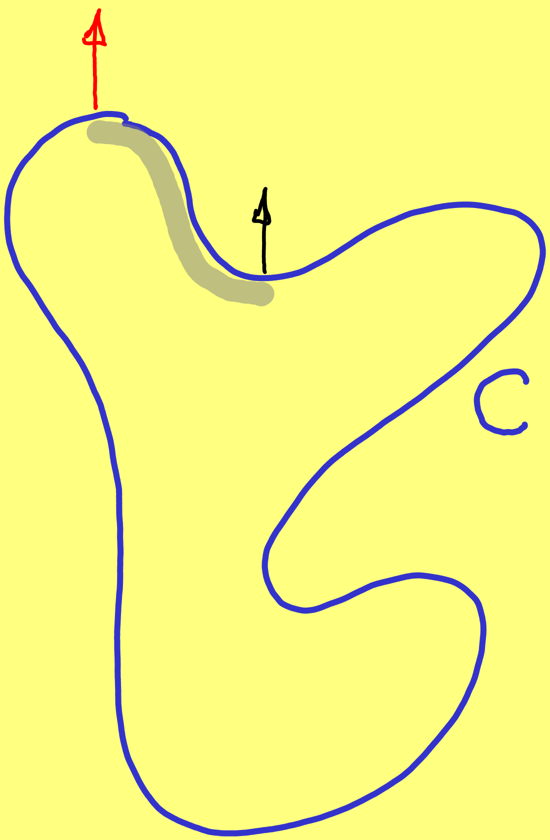
FÁRY: $L(C) \leq K(C)$.

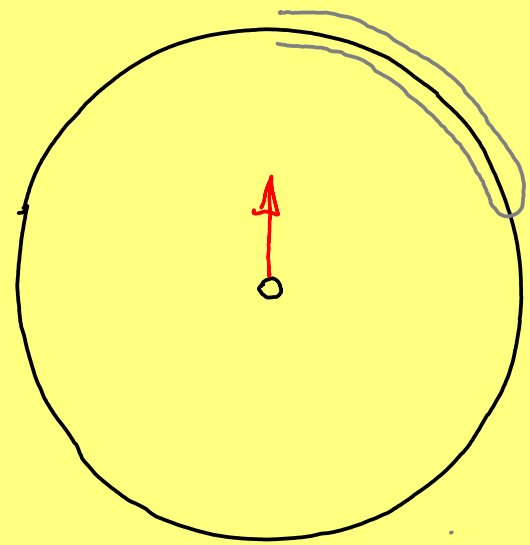
II.3 Total Curvature



$$K(C) = \int_{x \in C} |\kappa(x)|\, dx$$

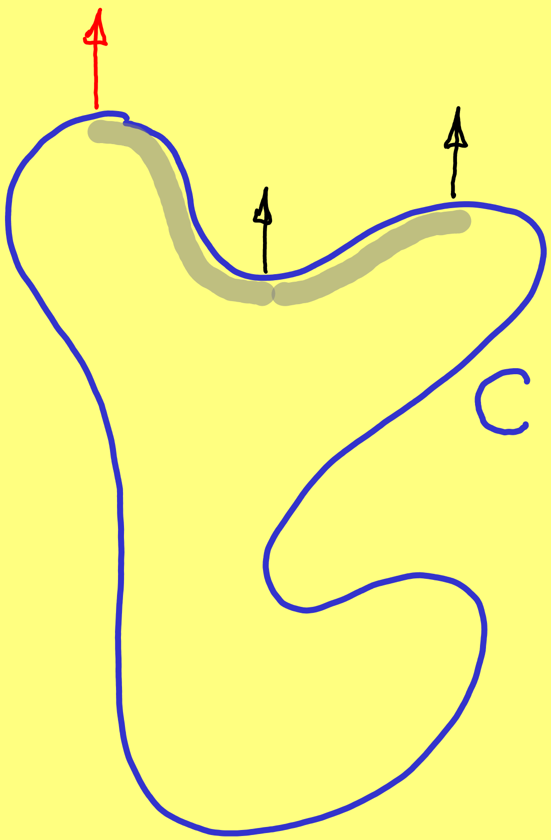
II.3 Total Curvature

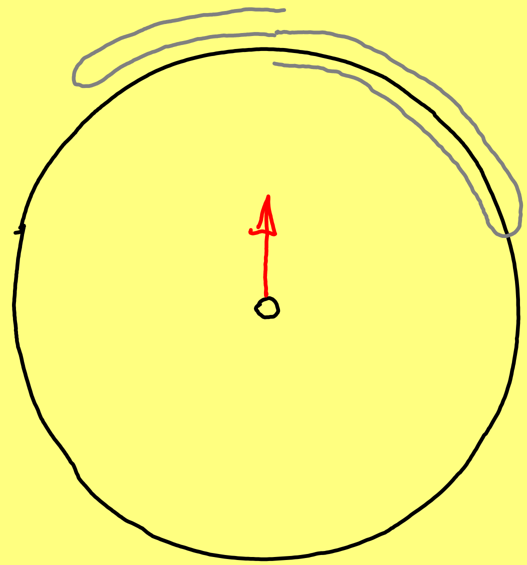




$$K(C) = \int_{x \in C} |\kappa(x)| \, dx$$

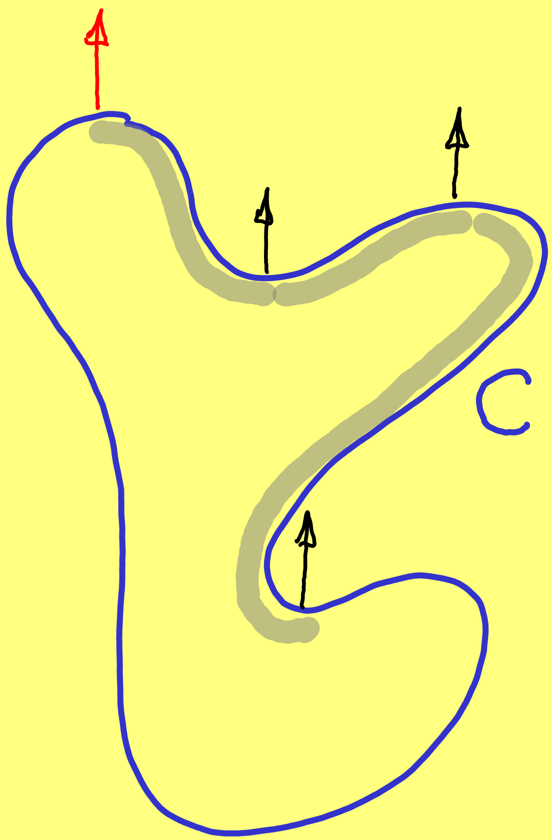
II.3 Total Curvature

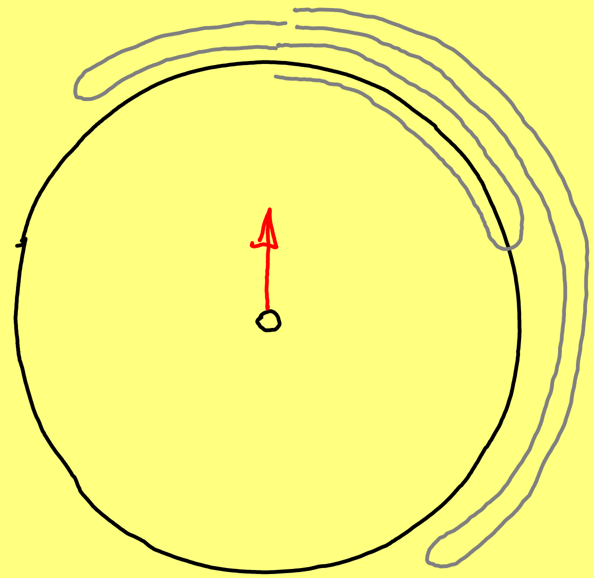




$$K(C) = \int_{x \in C} |\kappa(x)|\, dx$$

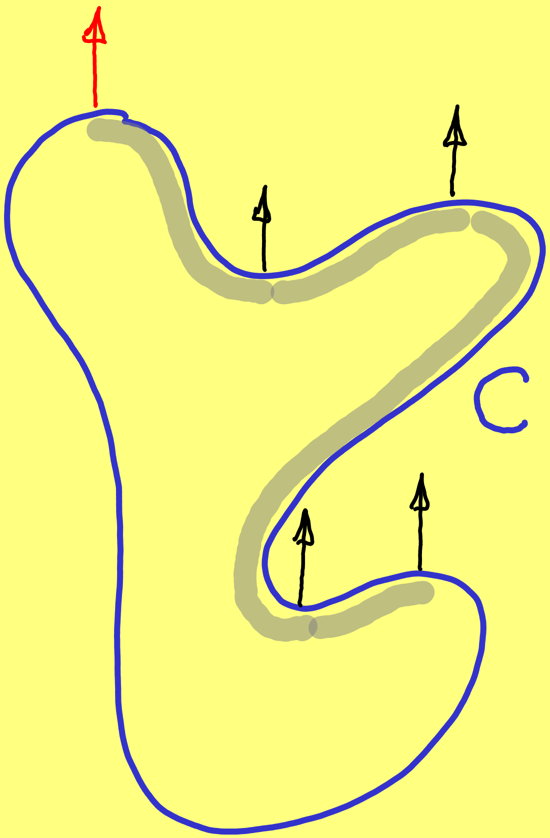
II.3 TOTAL CURVATURE

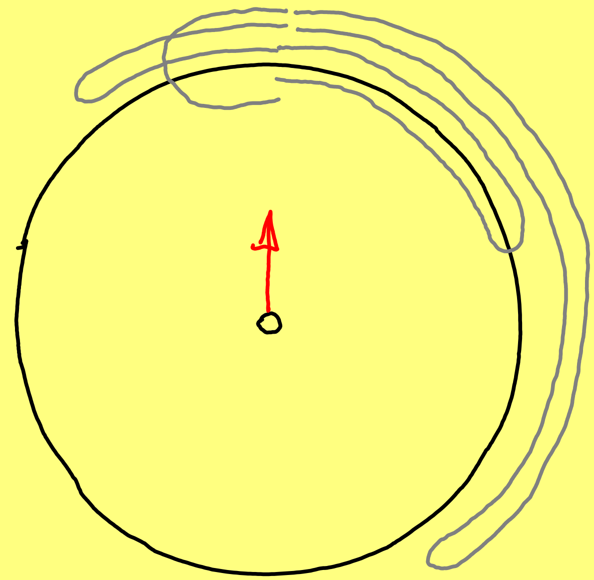




$$K(C) = \int_{x \in C} |\kappa(x)|\, dx$$

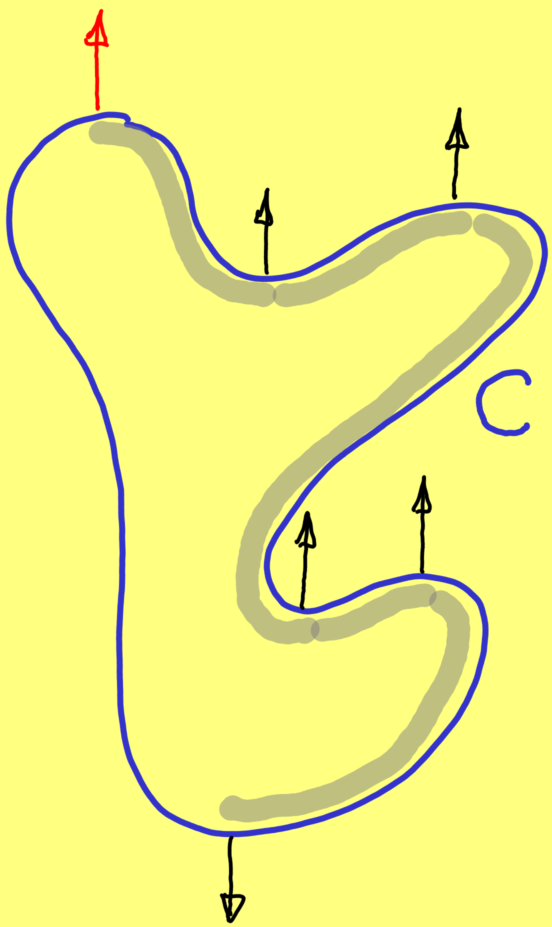
II.3 Total Curvature

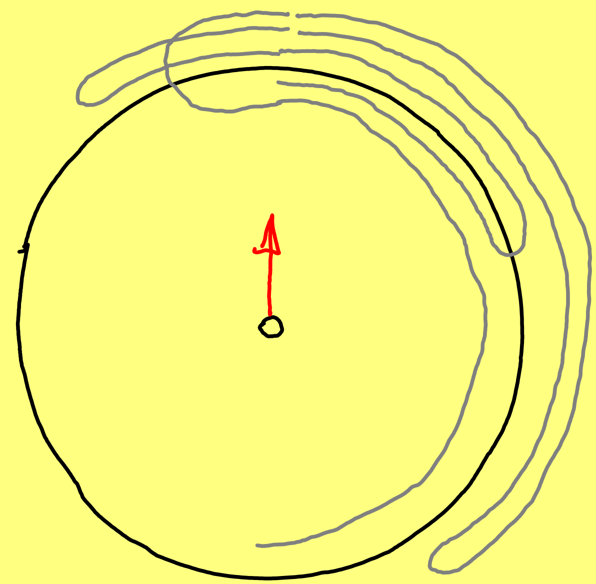




$$K(C) = \int_{x \in C} |\kappa(x)| \, dx$$

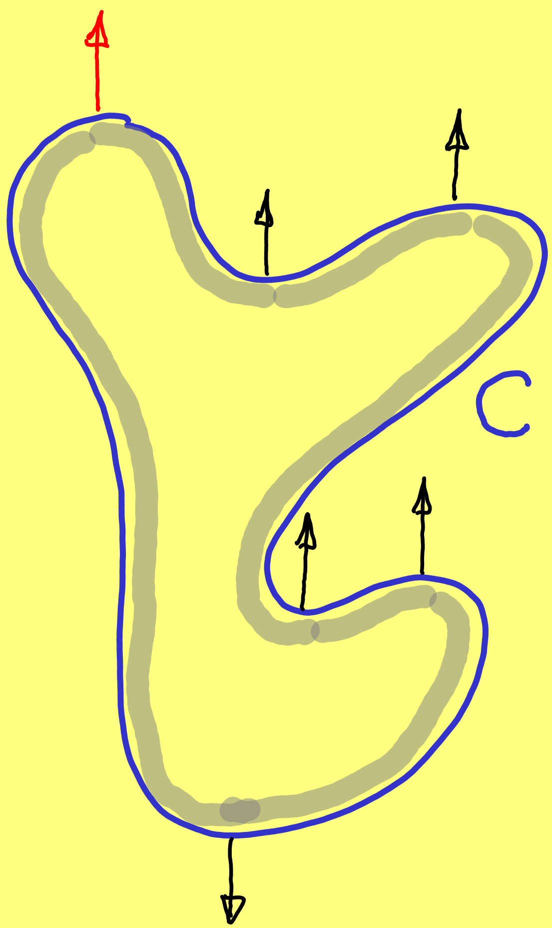
II.3 Total Curvature

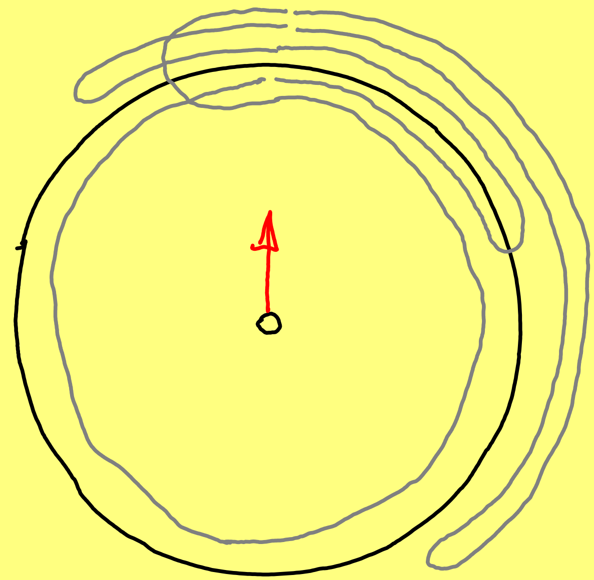




$$K(C) = \int_{x \in C} |\kappa(x)|\, dx$$

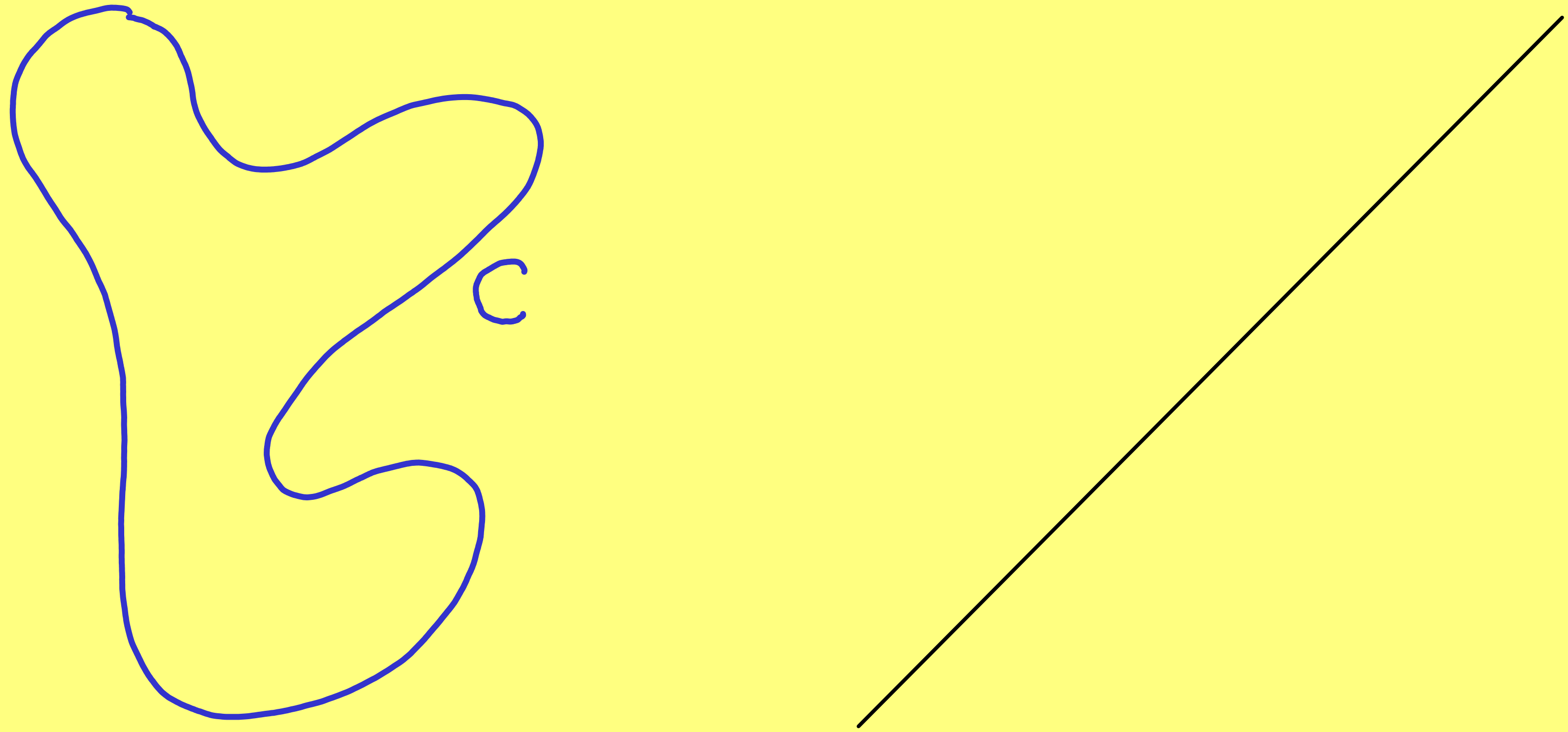
II.3 Total Curvature



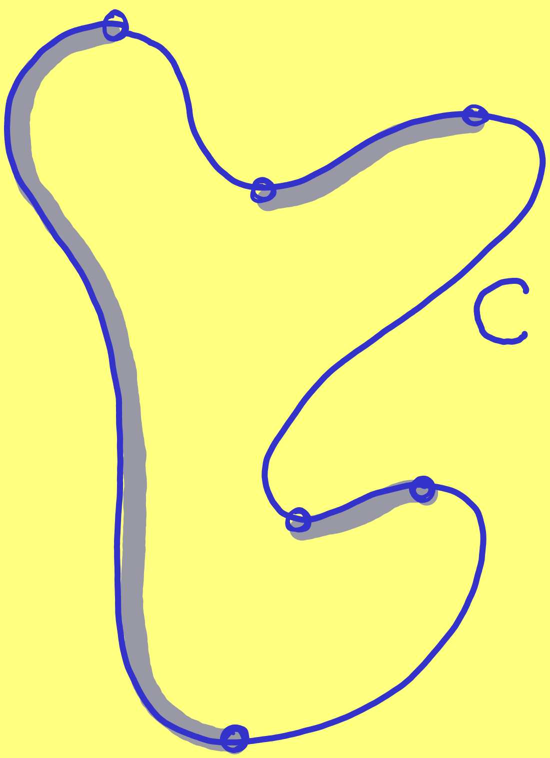


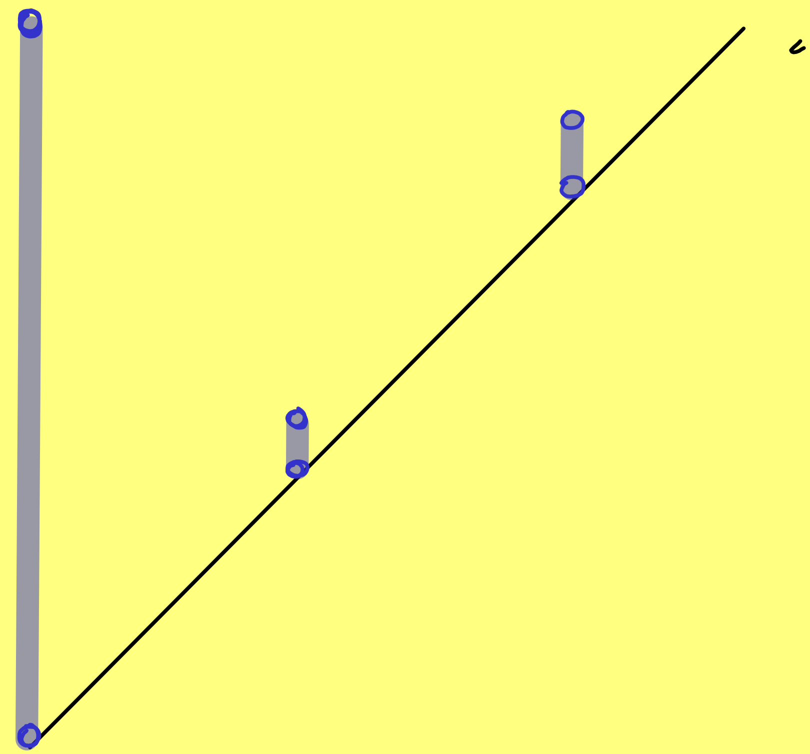
$$K(C) = \int_{x \in C} |\kappa(x)| \, dx$$

II.4 Generalized Fáry Theorem

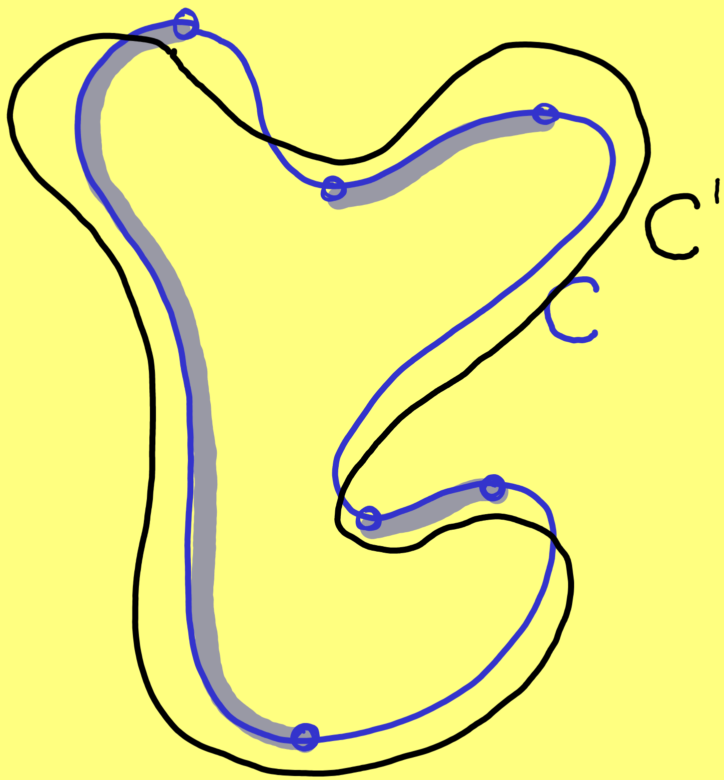


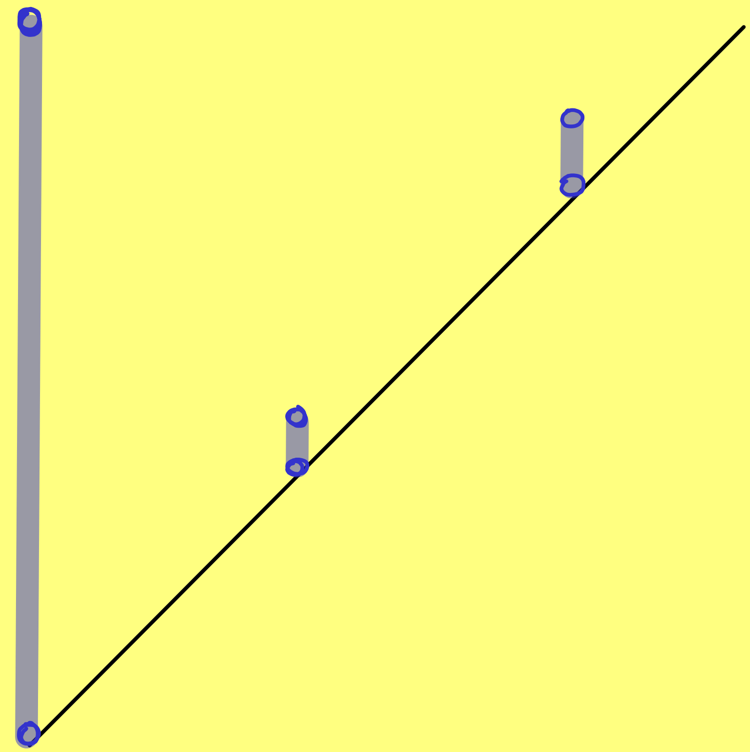
II.4 Generalized Fáry Theorem



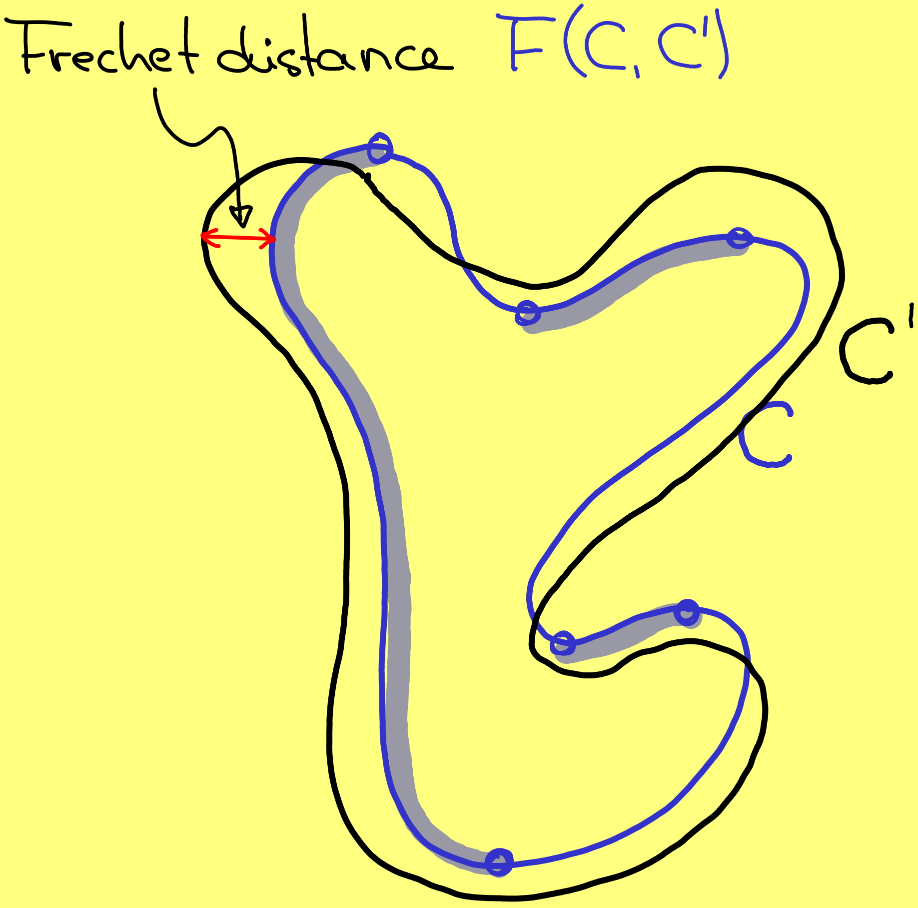


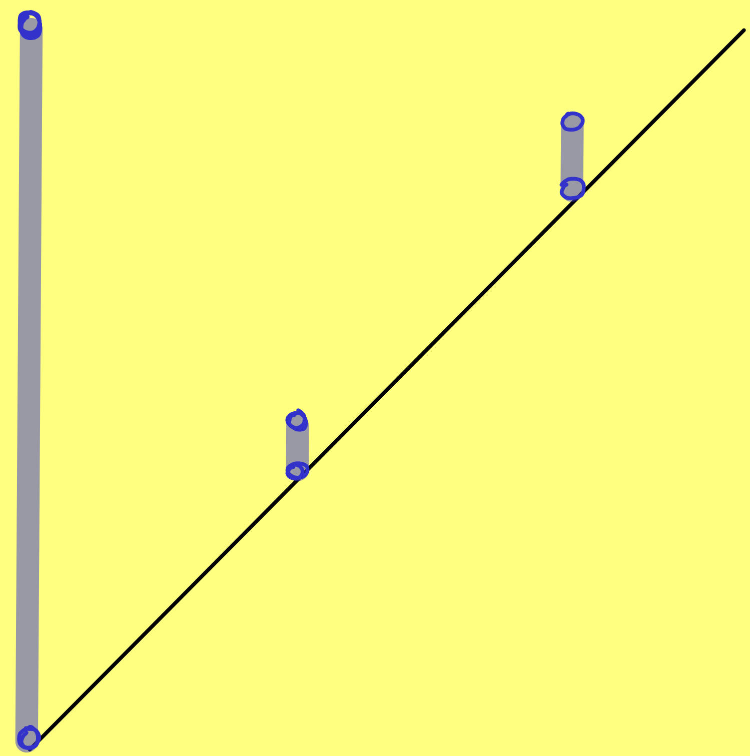
II.4 Generalized Fáry Theorem



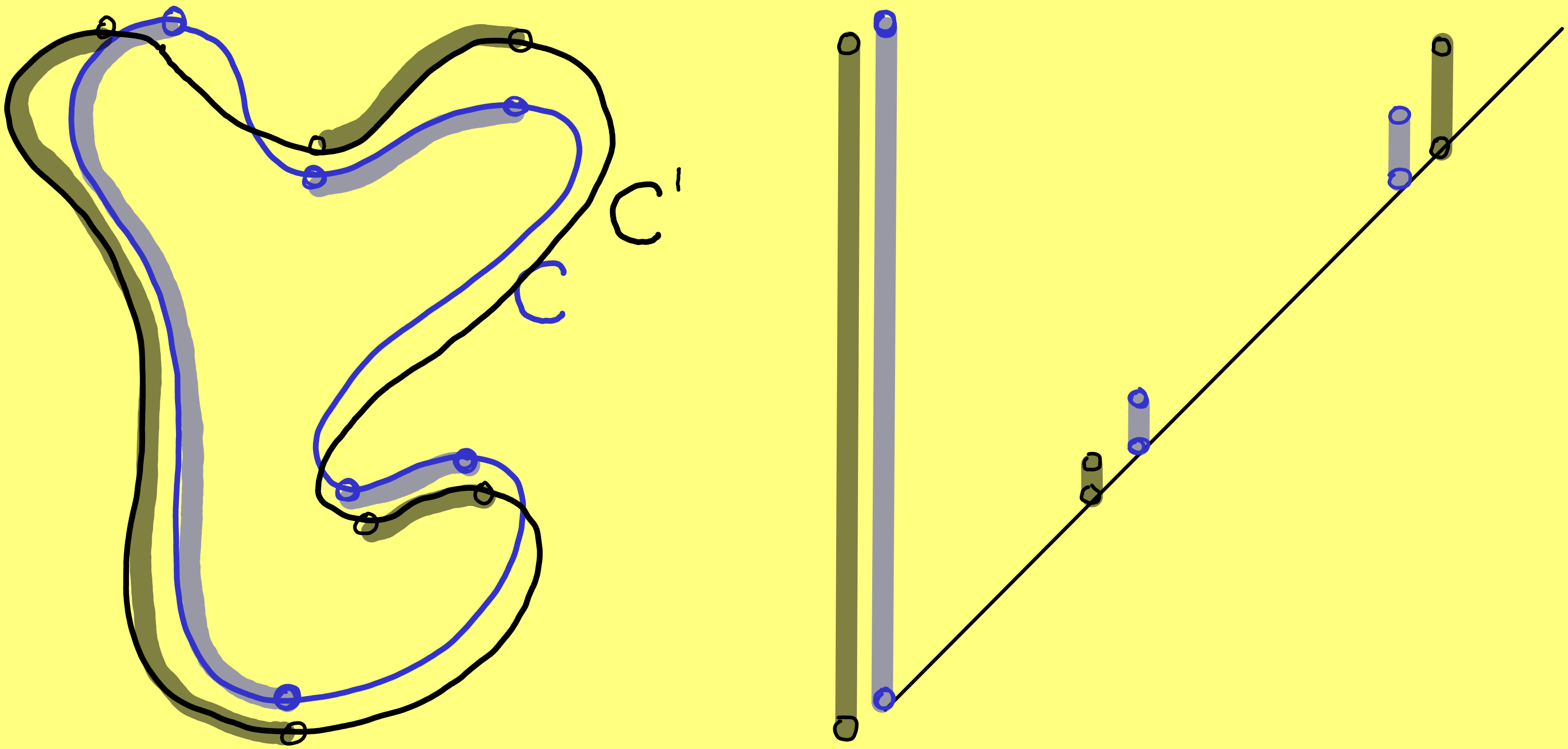


II.4 Generalized Fáry Theorem

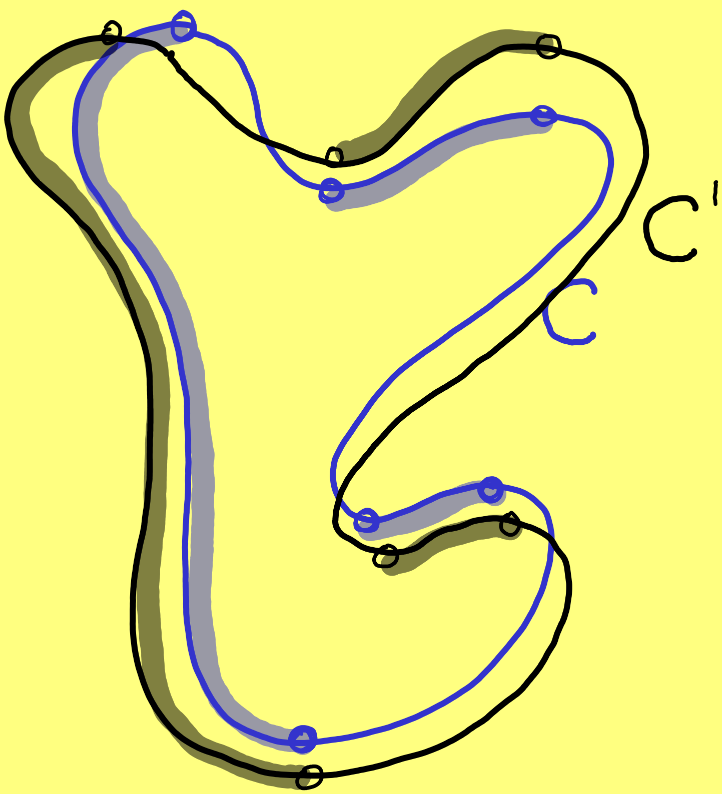


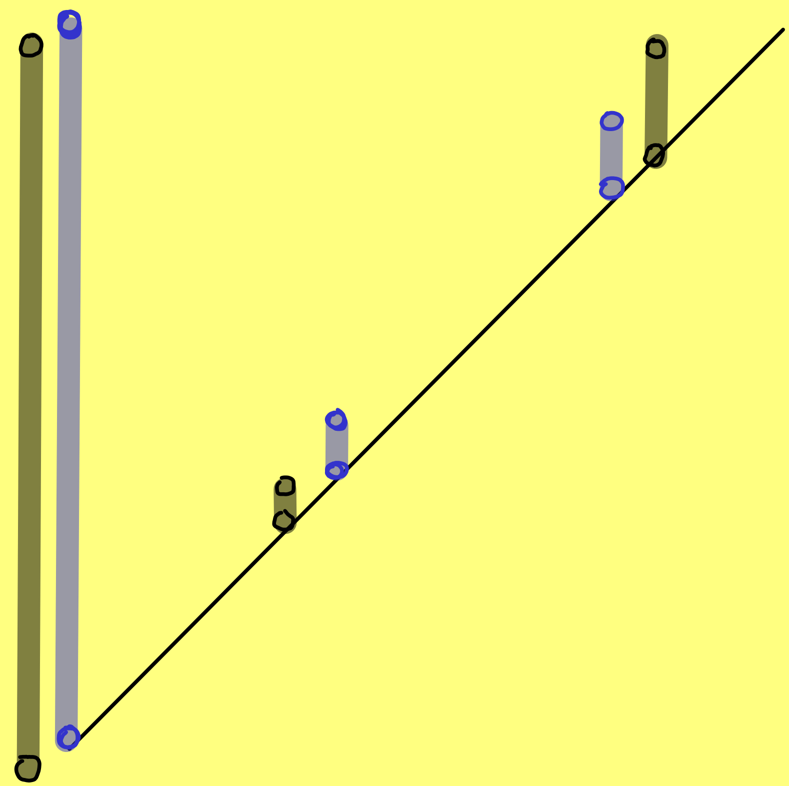


II.4 Generalized Fáry Theorem



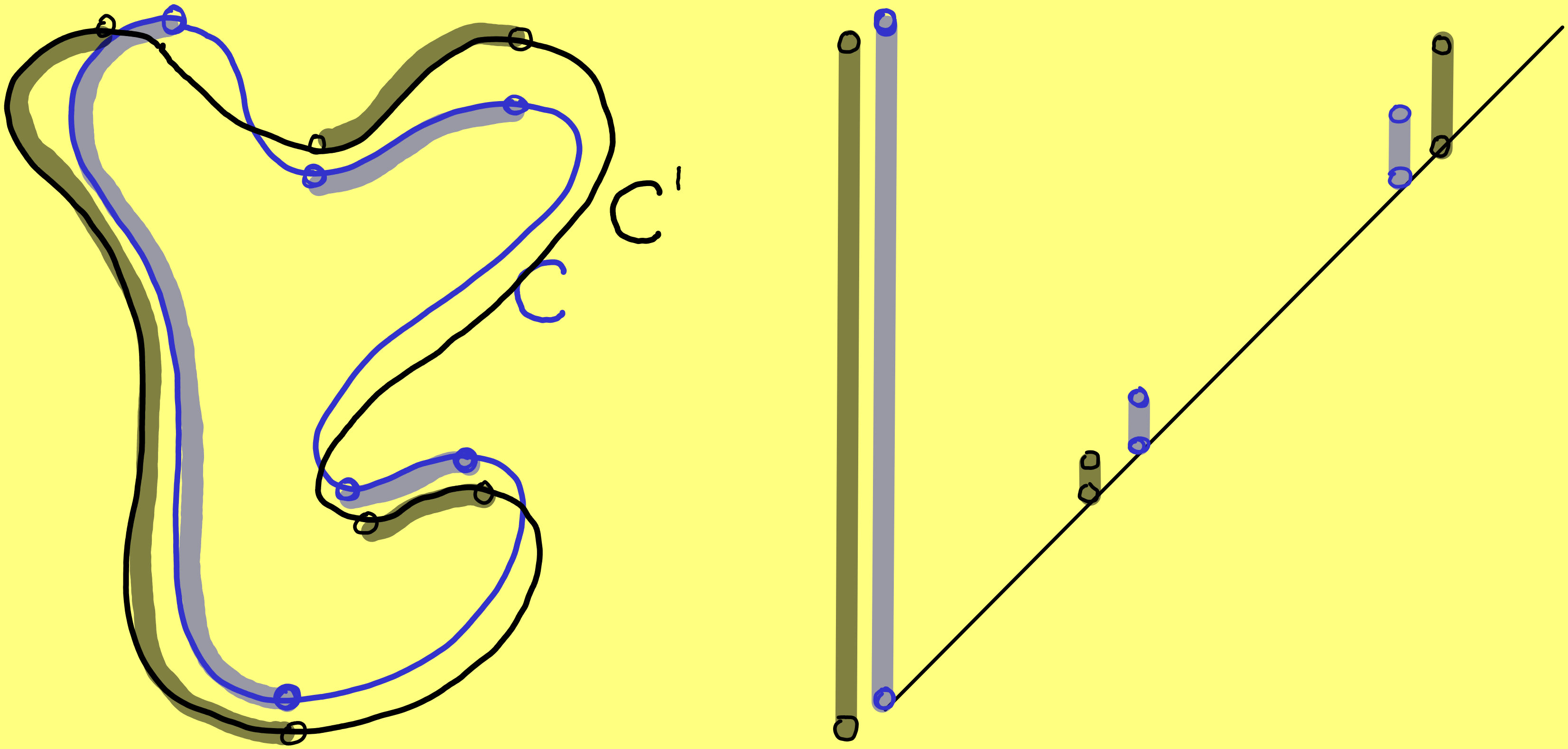
II.4 Generalized Fáry Theorem





Thm. $|L(C) - L(C')| \leq (K(C) + K(C') - 2\pi)\, F(C, C')$.

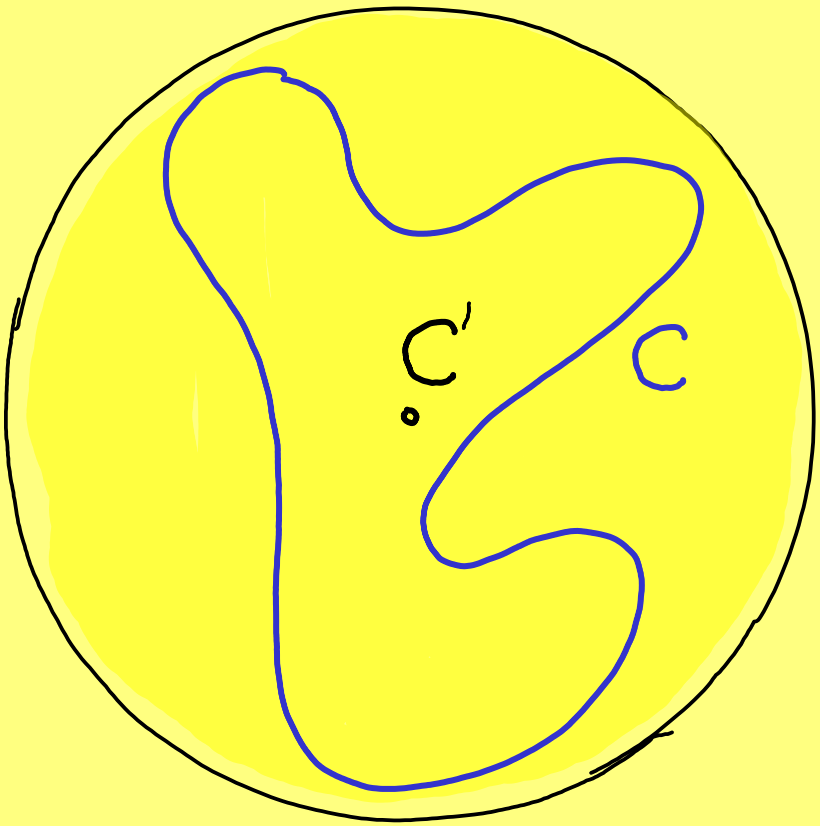
II.4 Generalized Fáry Theorem



2006 Cohen-Steiner, E

Thm. $|L(C) - L(C')| \leq (K(C) + K(C') - 2\pi)\, F(C, C')$.

II.4 Generalized Fáry Theorem



Fáry: $L(C) \le K(C)$.

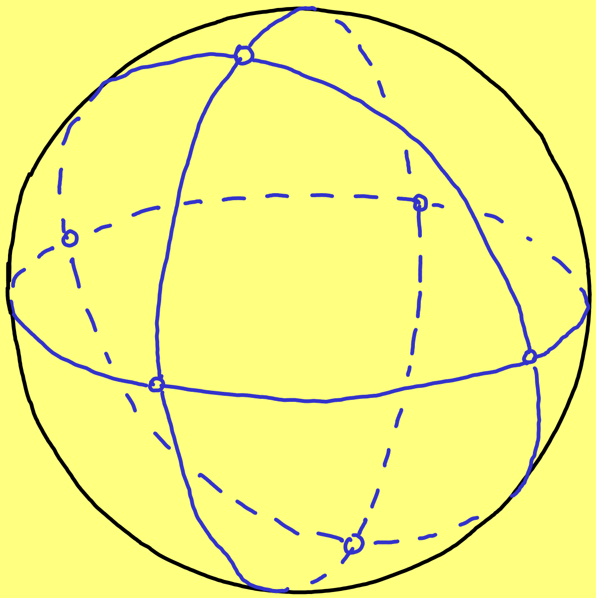
Thm. $|L(C) - L(C')| \le (K(C) + K(C') - 2\pi)\, F(C, C')$.

I Persistence

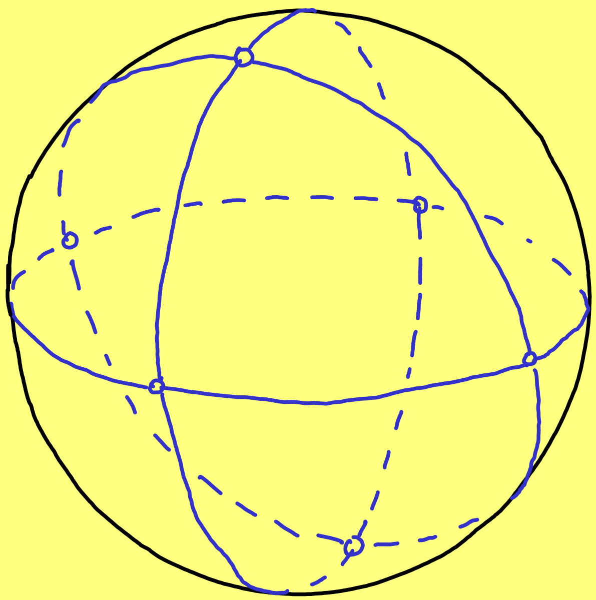
II Curves

III Somites

III.1 Size and Growth



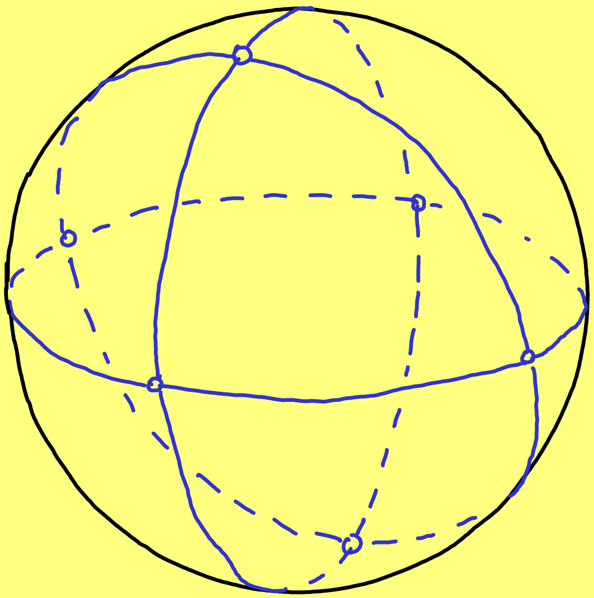
III.1 Size and Growth



mesh (K) = max. diameter of a simplex

size (K) = number of simplices

III.1 Size and Growth

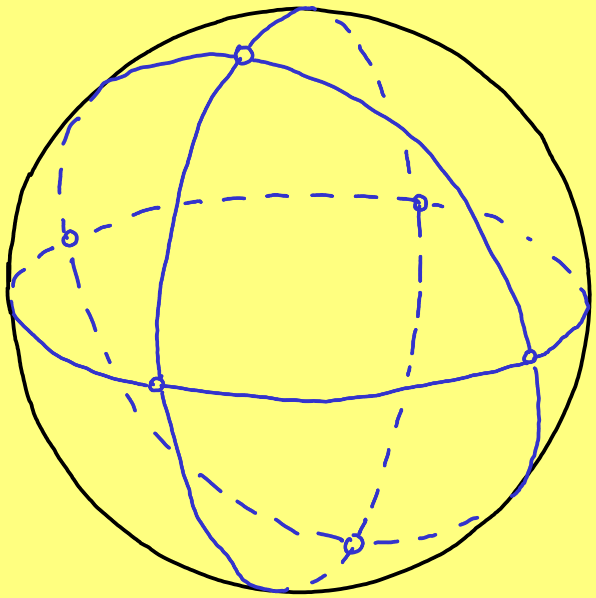


mesh (K) = max. diameter of a simplex

size (K) = number of simplices

$$N(r) = \min_{\text{mesh}(K) \le r} \text{size}(K)$$

III.1 Size and Growth



mesh (K) = max. diameter of a simplex

size (K) = number of simplices

$$N(r) = \min_{\text{mesh}(K) \le r} \text{size}(K)$$

The triangulation grows polynomially if $\exists$ constants C, M s.t.

$$N(r) \le C/r^M.$$

III.2 Total Persistence

The q-th total persistence of f is

$$\mathrm{Pers}_q(f) = \sum_p \sum_x \mathrm{pers}(x)^q$$

III.2 Total Persistence

The q-th total persistence of f is

$$\mathrm{Pers}_q(f) = \sum_p \sum_x \mathrm{pers}(x)^q$$

f Lipschitz and $N(r) \le C/r^M$

$\Longrightarrow$ $\mathrm{Pers}_q(f) \le \mathrm{const}$ for all $q > M$

III.2 Total Persistence

TP-Stab. Thm. Let $\mathbb{X}$ be a compact metric space with polynomially growing triangulation and $f, g : \mathbb{X} \to \mathbb{R}$ Lipschitz functions. Then

$$\mathrm{Pers}_q(f) - \mathrm{Pers}_q(g) \leq \mathrm{const} \cdot \|f-g\|_\infty$$

for every $q > M+1$.

III.2 Total Persistence

TP-STAB. THM. Let $\mathbb{X}$ be a compact metric space with polynomially growing triangulation and $f, g : \mathbb{X} \to \mathbb{R}$ Lipschitz functions. Then

$$\mathrm{Pers}_q(f) - \mathrm{Pers}_q(g) \le \mathrm{const} \cdot \|f-g\|_\infty$$

for every $q > M+1$.

2009 Cohen-Steiner, E., Harer, Mileyko

III.3 SIMPLIFICATION

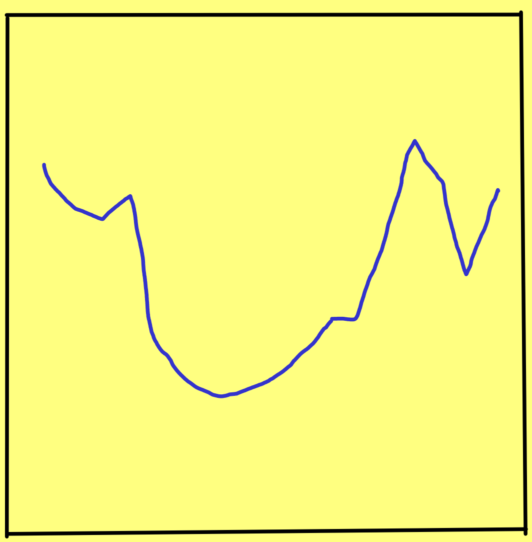
normalized to $\mathrm{amp}(f) = \max_x f(x) - \min_x f(x) = 1$

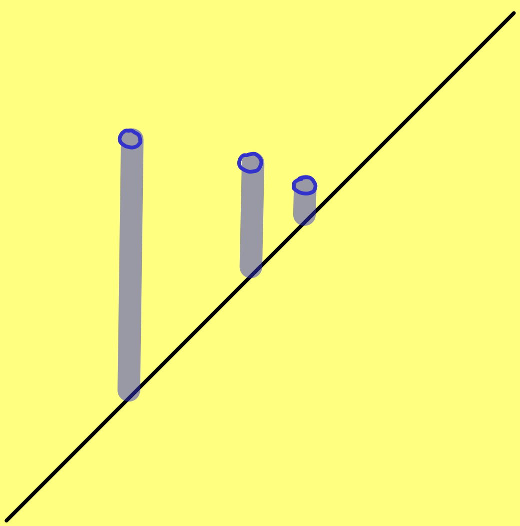
III.3 Simplification

An ε-simplification of f is a function $f_\varepsilon : \mathbb{X} \to \mathbb{R}$ with $\|f - f_\varepsilon\|_\infty \le \varepsilon$ whose diagrams $\mathrm{Dgm}_p(f_\varepsilon)$ are same as $\mathrm{Dgm}_p(f)$ without points of persistence at most ε.

III.3 Simplification

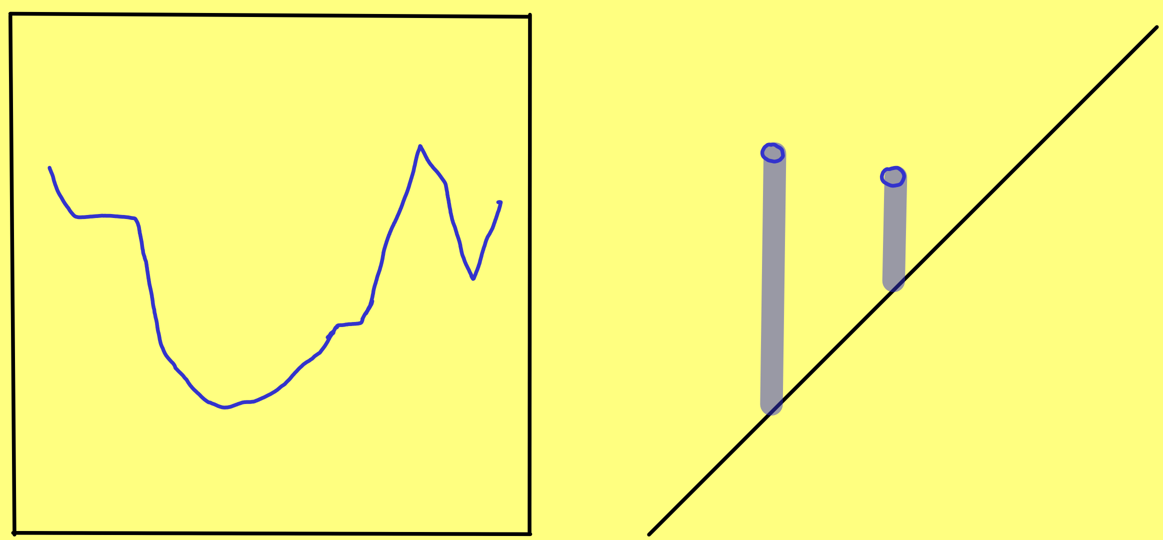
An ε-simplification of f is a function $f_\varepsilon : \mathbb{X} \to \mathbb{R}$ with $\| f - f_\varepsilon \|_\infty \leq \varepsilon$ whose diagrams $\mathrm{Dgm}_p(f_\varepsilon)$ are same as $\mathrm{Dgm}_p(f)$ without points of persistence at most ε.





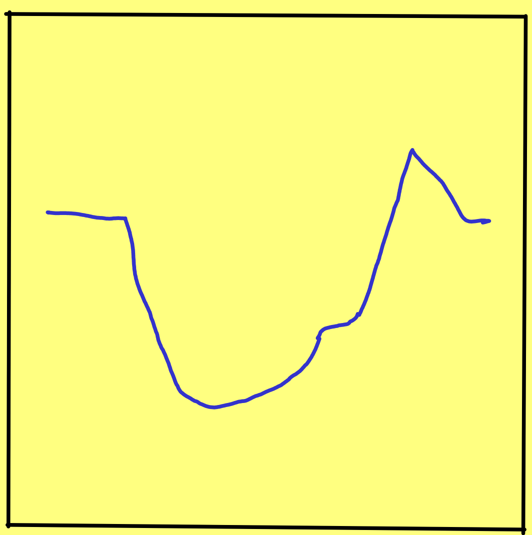
III.3 SIMPLIFICATION

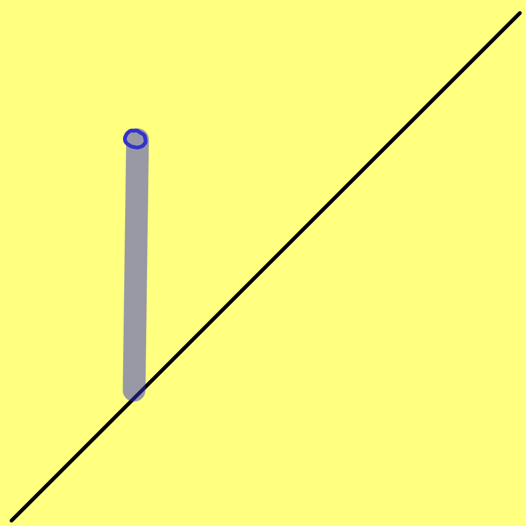
An ε-simplification of f is a function $f_\varepsilon : \mathbb{X} \to \mathbb{R}$ with $\|f - f_\varepsilon\|_\infty \leq \varepsilon$ whose diagrams $\mathrm{Dgm}_p(f_\varepsilon)$ are same as $\mathrm{Dgm}_p(f)$ without points of persistence at most ε.



III.3 Simplification

An ε-simplification of f is a function $f_\varepsilon : \mathbb{X} \to \mathbb{R}$ with $\|f - f_\varepsilon\|_\infty \le \varepsilon$ whose diagrams $\mathrm{Dgm}_p(f_\varepsilon)$ are same as $\mathrm{Dgm}_p(f)$ without points of persistence at most ε.





III.3 Simplification

$f : S^1 \to \mathbb{R}$

$M_0(f) = \frac{1}{2}$ #crit. pts.

III.3 SIMPLIFICATION

$f : \mathbb{S}^1 \to \mathbb{R}$

$$M_0(f) = \frac{1}{2} \# \text{crit. pts.}$$

$$M_1(f) = \int_{\varepsilon=0}^{1} M_0(f_\varepsilon)\, d\varepsilon$$

III.3 SIMPLIFICATION

$f : S^1 \to \mathbb{R}$

$$M_0(f) = \tfrac{1}{2} \, \#\text{crit. pts.}$$

$$M_1(f) = \int_{\varepsilon=0}^{1} M_0(f_\varepsilon)\, d\varepsilon$$

$$M_2(f) = \int_{\varepsilon=0}^{1} M_1(f_\varepsilon)\, d\varepsilon$$

etc.

III.3 SIMPLIFICATION

$f : \mathbb{S}^1 \to \mathbb{R}$

$$M_0(f) = \frac{1}{2} \#\text{crit. pts.} = \mathrm{Pers}_0(f)$$

$$M_1(f) = \int_{\varepsilon=0}^{1} M_0(f_\varepsilon)\, d\varepsilon = \mathrm{Pers}_1(f)$$

$$M_2(f) = \int_{\varepsilon=0}^{1} M_1(f_\varepsilon)\, d\varepsilon = \mathrm{Pers}_2(f)$$

etc.

III.3 SIMPLIFICATION

$f : \mathbb{S}^1 \to \mathbb{R}$

not stable

$$M_0(f) = \frac{1}{2} \#\text{crit. pts.} = \mathrm{Pers}_0(f)$$

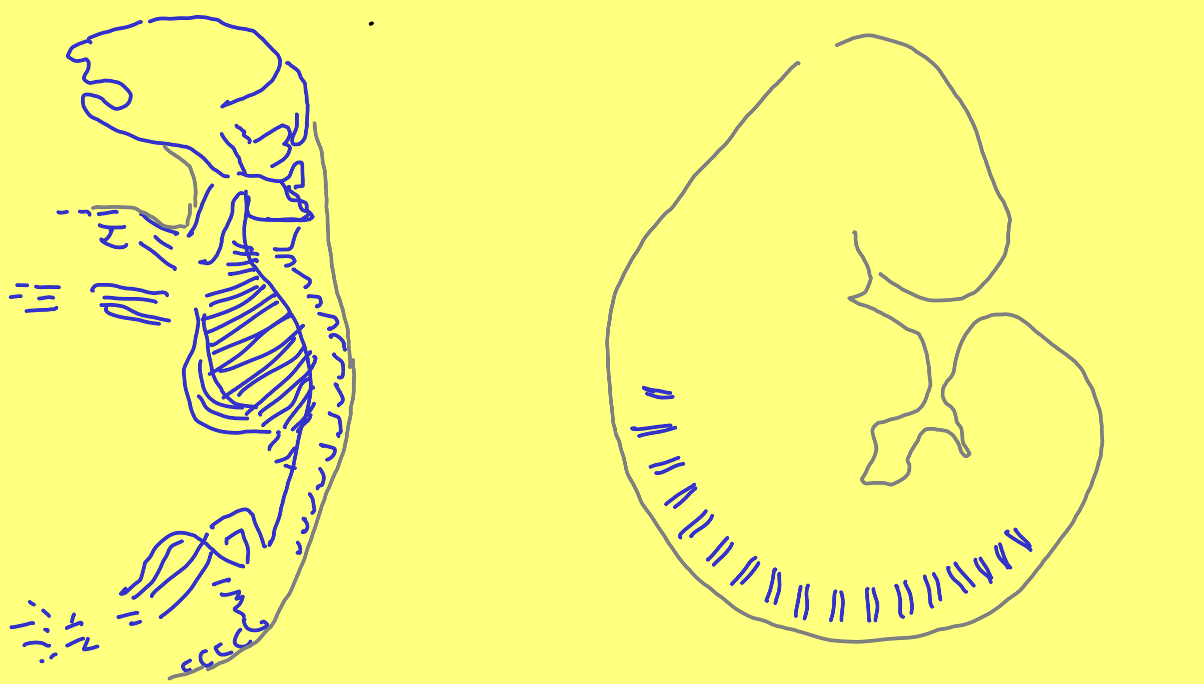
$$M_1(f) = \int_{\varepsilon=0}^{1} M_0(f_\varepsilon)\, d\varepsilon = \mathrm{Pers}_1(f)$$

stable

$$M_2(f) = \int_{\varepsilon=0}^{1} M_1(f_\varepsilon)\, d\varepsilon = \mathrm{Pers}_2(f)$$

etc.

III.4 Rhythmic Gene Expression

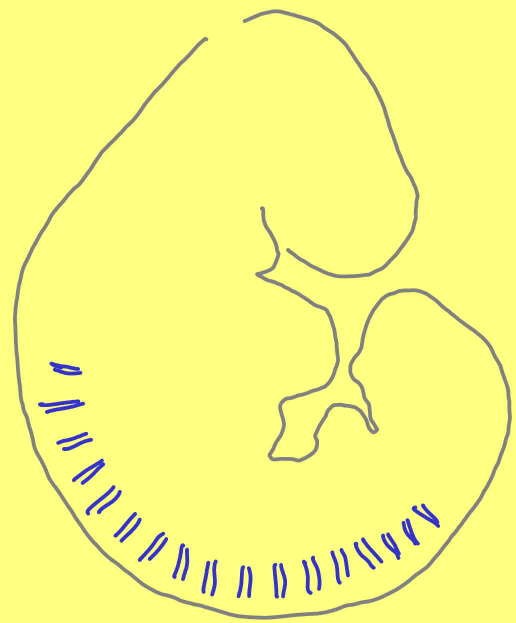


somite development is a rythmic process

III.4 RHYTHMIC GENE EXPRESSION

adult mouse

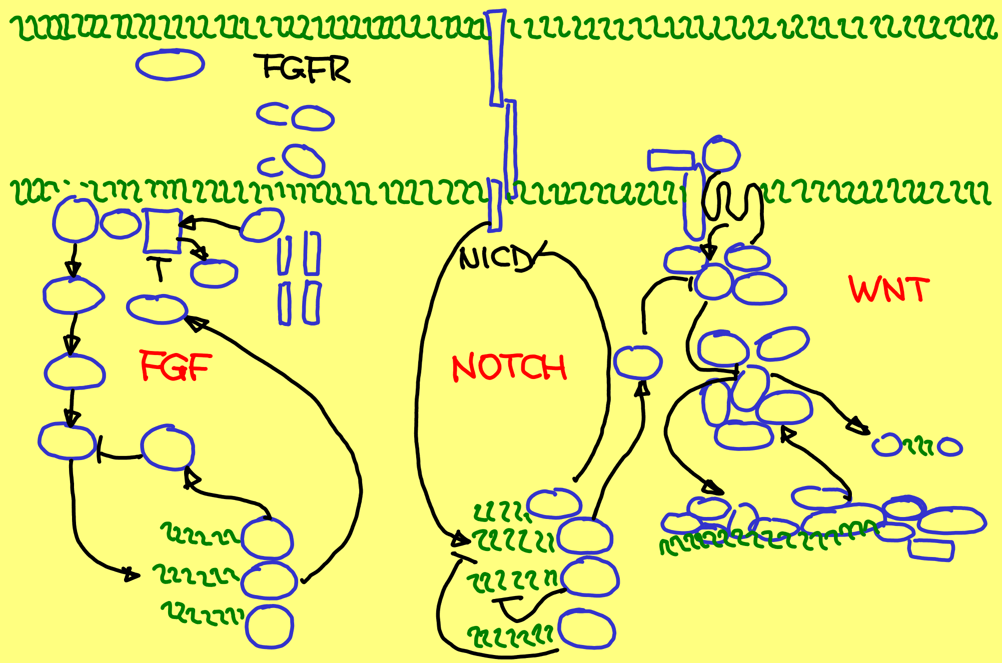
mouse embryo



somite development is a rythmic process

Pourqie Lab, Stowers

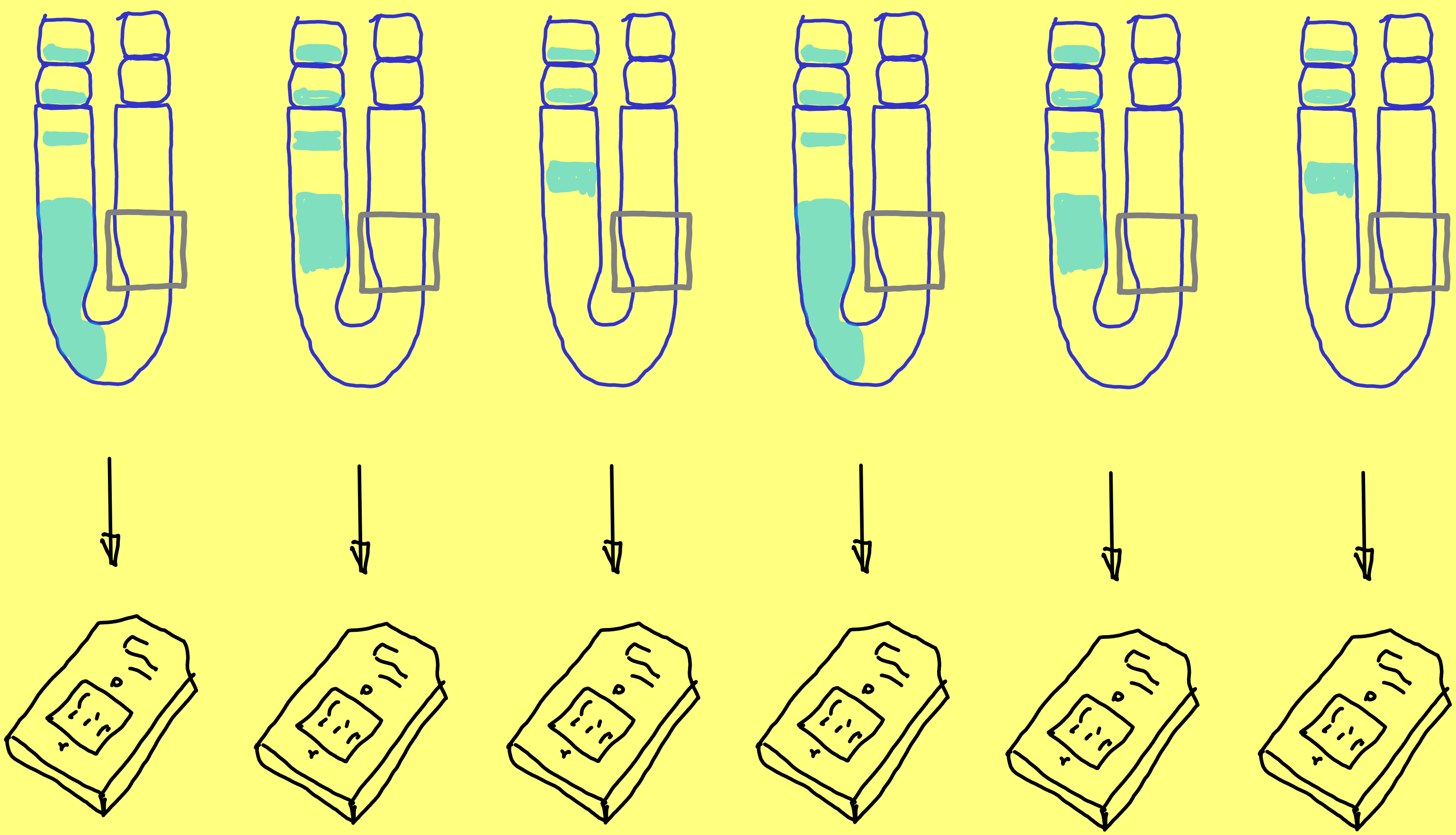
III.4 Rhythmic Gene Expression



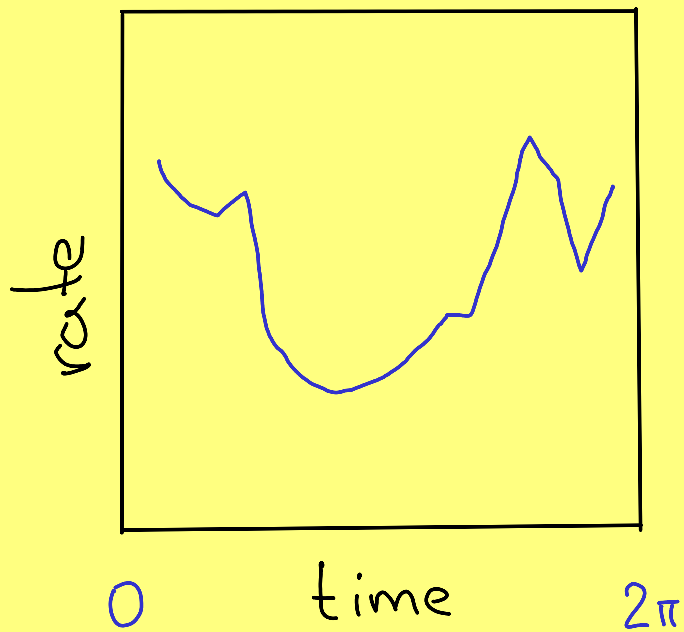
the proposed clock

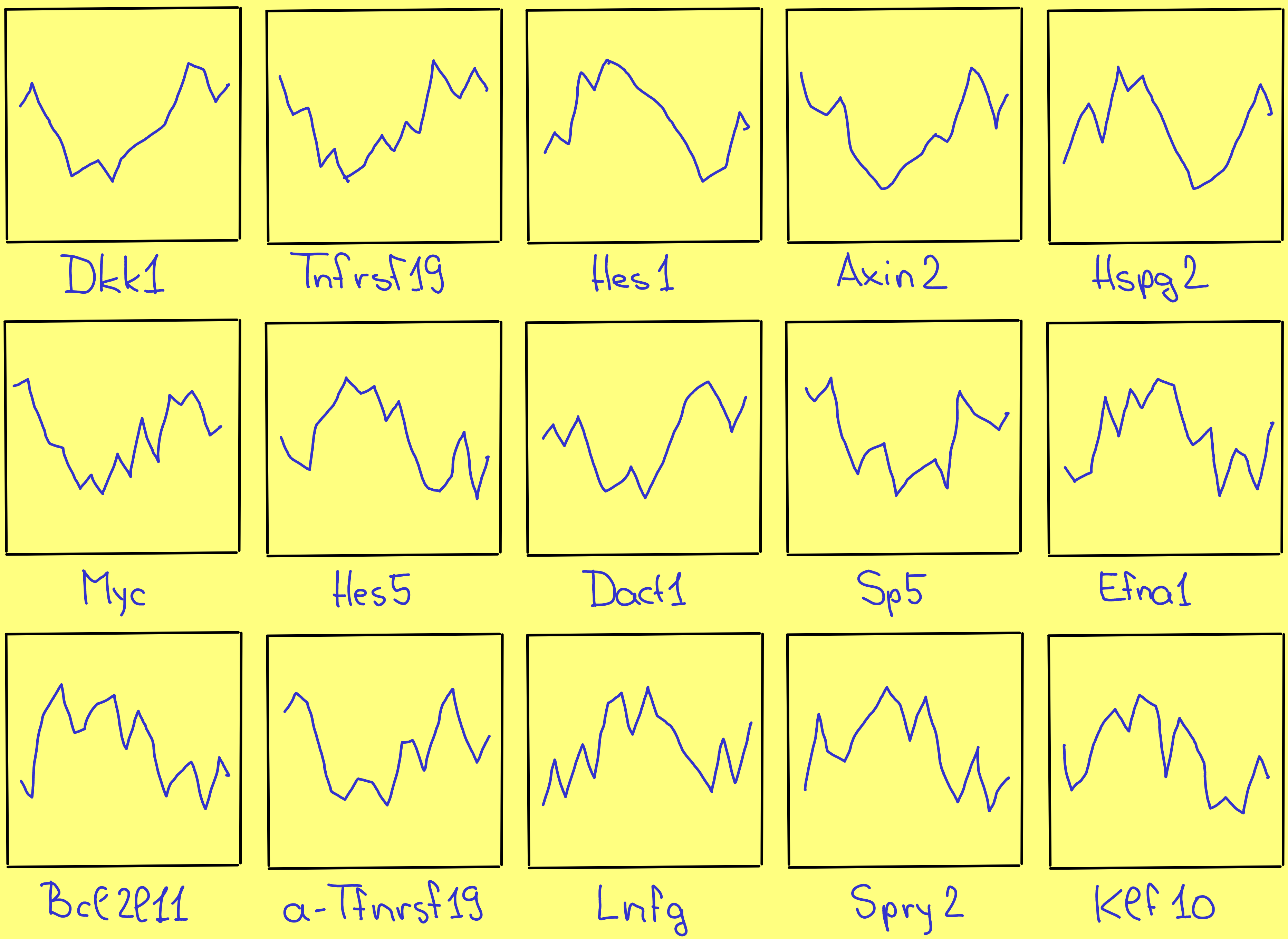
III.4 Rhythmic Gene Expression

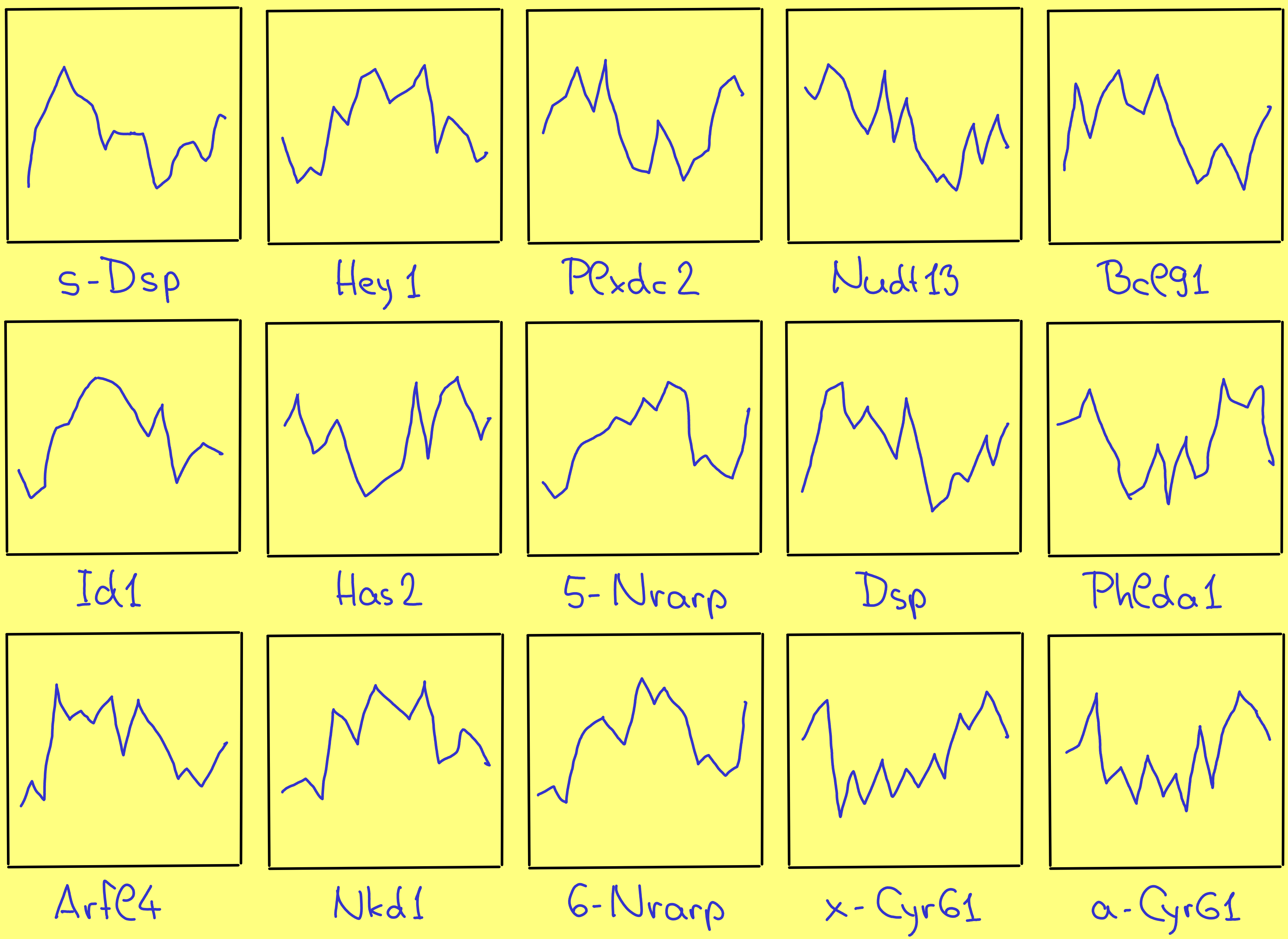
... time series microarray analysis



III.4 Rhythmic Gene Expression



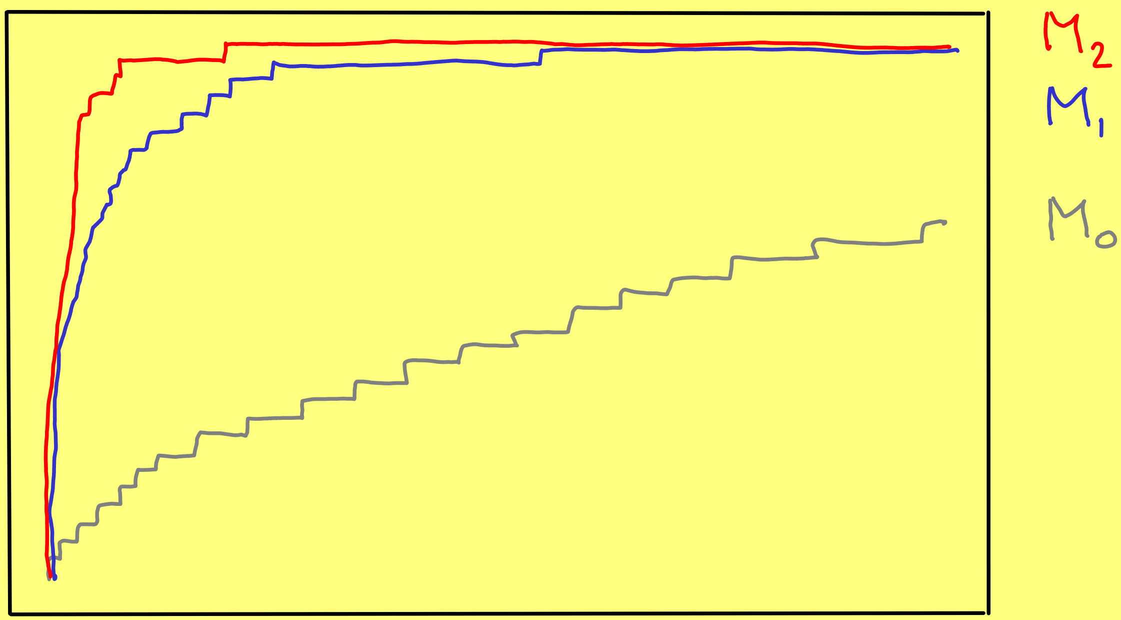

Dkk1
Tnfrsf19
Hes1
Axin2
Hspg2
Myc
Hes5
Dact1
Sp5
Efna1
Bcl2l11
a-Tfnrsf19
Lnfg
Spry2
Klf10


s-Dsp
Hey1
Plxdc2
Nudt13
Bcl9l
Id1
Has2
5-Nrarp
Dsp
Phlda1
Arl4
Nkd1
6-Nrarp
x-Cyr61
a-Cyr61

III.4 Rhythmic Gene Expression

yield of 30 biologically confirmed genes

ROC curves



	M_0	M_1	M_2	M_3	M_4	max
Area	4.42	9.75	10.18	10.19	10.15	10.50

THANK YOU