From Macro to Micro with Möbius: A compositional approach to higher-order structure

Imperial, Centre for Complexity Science – April 2024

Abel Jansma
@Abelaer

Max Planck Institute for Mathematics in the Sciences
School of Informatics, University of Edinburgh

MAX-PLANCK-INSTITUT
FÜR MATHEMATIK
IN DEN NATURWISSENSCHAFTEN

Introduction

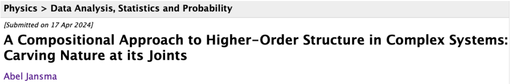
- MSc Theoretical Physics @ Uni of Amsterdam
- PhD Biomedical AI @ Uni of Edinburgh
- Higher-order information theory: genes and Ising models
- Met Fernando and Pedro in Dresden $\implies$ now with Jürgen Jost & Bernd Sturmfels @ MPI MiS, Leipzig

Overview

GOAL: Higher-order structure $\iff$ System decomposition.

1. Decomposition in the forward and inverse problem
2. Möbius inversions: maths and philosophy
3. Examples: information theory, biology, physics, game theory, AI
4. Summary and outlook

(Based on arXiv:2404.14423)

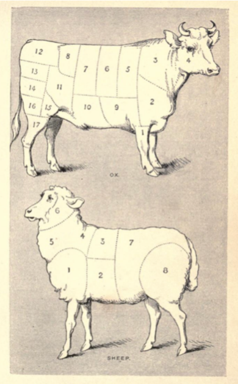


Physics > Data Analysis, Statistics and Probability

[Submitted on 17 Apr 2024]

A Compositional Approach to Higher-Order Structure in Complex Systems: Carving Nature at its Joints

Abel Jansma



Handbook of practical cookery
M. Dods, 1886

Part 1: (De-)Composition

Genotype-phenotype mapping

- Say the height H of a person is determined by the effect h two genetic variants:

$$H(\emptyset) = h(\emptyset) \tag{1}$$

$$H(\{g_1\}) = h(\emptyset) + h(\{g_1\}) \tag{2}$$

$$H(\{g_2\}) = h(\emptyset) + h(\{g_2\}) \tag{3}$$

$$H(\{g_1, g_2\}) = h(\emptyset) + h(\{g_1\}) + h(\{g_2\}) + h(\{g_1, g_2\}) \tag{4}$$

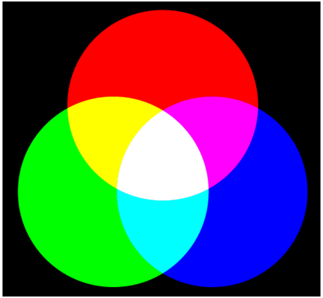
  Then the genetic effects can be estimated from observations of people's heights:

$$h(\{g_1\}) = H(\{g_1\}) - H(\emptyset) \tag{5}$$

$$h(\{g_1, g_2\}) = H(\{g_1, g_2\}) - H(\{g_1\}) - H(\{g_2\}) + H(\emptyset) \tag{6}$$

Colour theory

- **Intuition:** Magenta has some *redness*, *blueness*, (*blackness*), and an interaction between *redness* and *blueness*.
- Additive colour mixing:



$$Red = I_{Red} + I_{Black}$$

$$Magenta = I_{Red} + I_{Blue} + I_{Red,Blue} + I_{Black}$$

$$White = I_{Red} + I_{Green} + I_{Blue} + I_{Red,Blue} + I_{Green,Blue} + I_{Red,Green} + I_{Red,Green,Blue} + I_{Black}$$

Inverse problem of defining colour interactions:

$$I_{Red,Blue} = I_{Magenta} = Magenta - Red - Blue + Black$$

- The interaction between red and blue is what's in magenta, but not in red or blue.
- **Claim:** This is a very general construction in complex systems theory.

- **General construction:** A macroscopic quantity Q of a system S is a sum over microscopic contributions q of *parts* of a decomposition $\mathcal{D}(S)$.

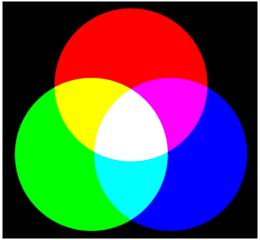
$$Q(S) = \sum_{t \in \mathcal{D}(S)} q(t) \tag{7}$$

Example: powerset decomposition

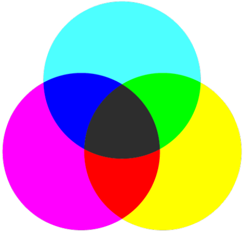
$$Q(S) = \sum_{t \in \mathcal{P}(S)} q(t) = \sum_{t \subseteq S} q(t) \tag{8}$$

- Forward problem: given q, find Q
- Inverse problem: given Q, find q.
- Can the sum (7) be inverted?
- **Claim:** Yes, the pair $(Q, \mathcal{D})$ uniquely defines the microscopic quantity q

Different decompositions



Additive colour mixing



Subtractive colour mixing

Part 2: Functions on Partial Orders

- How to invert a macro-micro decomposition
- (The mathsy part)

Algebra of functions on intervals

Definition

A **partial order** on a set P is a binary relation $\leq$, such that for all $a, b, c \in P$:

$$\text{Reflexivity: } a \leq a \tag{9}$$

$$\text{Transitivity: } a \leq b \text{ and } b \leq c \implies a \leq c \tag{10}$$

$$\text{Antisymmetry: } a \leq b \text{ and } b \leq a \iff a = b \tag{11}$$

- Consider: functions on intervals $[a, b] = \{x : a \leq x \leq b\}$
- **Incidence algebra** with elements $f : P \times P \to \mathbb{R}$ and multiplication $*$

$$(f * g)(a, b) = \sum_{x : a \leq x \leq b} f(a, x) g(x, b) \tag{12}$$

- (P should be *locally finite*)

Special elements of the incidence algebra

- **Incidence algebra:** elements $f\colon P \times P \to \mathbb{R}$ and $(f * g)(a, b) = \sum_{a \leq x \leq b} f(a, x) g(x, b)$

$$*\text{-identity: } \delta_P(a, b) = \begin{cases} 1 & \text{if } a = b \\ 0 & \text{otherwise} \end{cases} \qquad (\delta_P * f = f = f * \delta_P) \tag{13}$$

$$\text{Constant function: } \zeta_P(a, b) = \begin{cases} 1 & \text{if } a \leq b \\ 0 & \text{otherwise} \end{cases} \tag{14}$$

- For a function $f(x) := f(\hat{0}, x)$, 'integration' is $(f * \zeta_P)(\hat{0}, b) = \sum_{x \leq b} f(x)$.
- cf. $Q(S) = \sum_{t \in \mathcal{D}(S)} q(t) = (q * \zeta_{\mathcal{D}(S)})(\hat{0}, S)$
- **Claim:** ζ_P has a unique $*$-inverse that inverts the sum.

The Möbius Inversion Theorem

$$\text{Möbius function: } \mu_P(x, y) = \begin{cases} 1 & \text{if } x = y \\ -\sum_{z: x \leq z < y} \mu_P(x, z) & \text{if } x < y \\ 0 & \text{otherwise} \end{cases}$$

- μ_P is the $*$-inverse of ζ_P: $\quad \mu_P * \zeta_P = \delta_p = \zeta_P * \mu_P$

Theorem (Rota 1964)

Let $(S, \leq)$ be a locally finite poset and $a, b \in S$. Let $f : S \to \mathbb{R}$ be a function on S, and let μ_S be the Möbius function on S. Then the following two statements are equivalent:

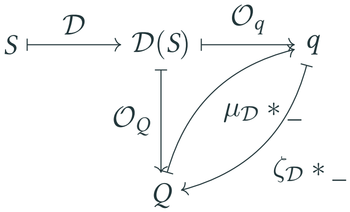
$$f(a) = \sum_{b \leq a} g(b) \iff g(a) = \sum_{b \leq a} \mu_S(b, a) f(b)$$

From Macro to Micro with Möbius

- Recall the decomposition $Q(S) = \sum_{t \in \mathcal{D}(S)} q(t)$
- If $\mathcal{D}(S)$ is a poset with max $\hat{1}$, then

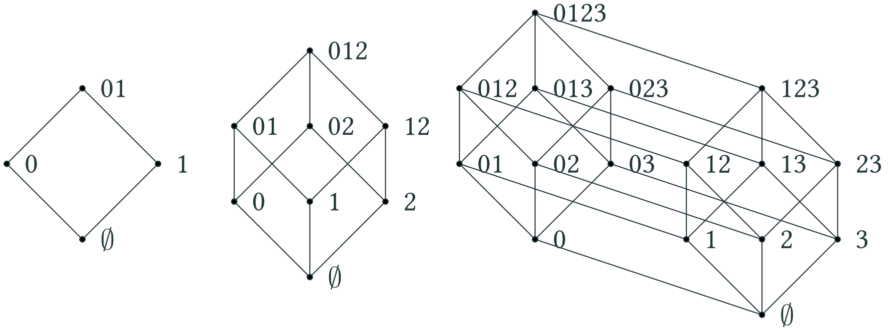
$$Q(S) = \sum_{t \leq \hat{1}} q(t) \qquad \iff \qquad q(t) = \sum_{t \leq \hat{1}} \mu_{\mathcal{D}(S)}(t, \hat{1}) Q(t)$$

- Estimate q: Macroscopic observations of Q, weighted by Möbius function.
- **The Möbius function of the decomposition solves the inverse problem.**



Example: Decomposing into the Power Set

- Power set $\mathcal{P}(S)$ with inclusion: poset $(\mathcal{P}(S), \subseteq)$
- The transitive reduction (Hasse diagram) forms a hypercube
- Very simple Möbius function: $\mu(t, s) = (-1)^{|s|-|t|}$
- e.g. $\mu(\{0\}, \{0, 1, 2\}) = (-1)^{3-1} = 1$



Example: Decomposing into the power set

- Say the height H of a person is determined by the effect h of two genetic variants:

$$H(\{g_1, g_2\}) = h(\emptyset) + h(\{g_1\}) + h(\{g_2\}) + h(\{g_1, g_2\}) \tag{15}$$

$$H(\{g_1, g_2\}) = \sum_{t \subseteq \{g_1, g_2\}} h(t) \tag{16}$$

$$\iff \quad h(\{g_1, g_2\}) = \sum_{t \subseteq \{g_1, g_2\}} \mu(t, \{g_1, g_2\}) H(t) \tag{17}$$

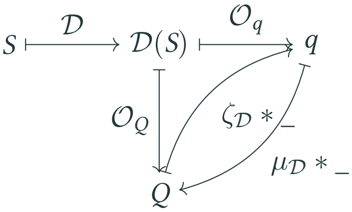
$$= H(\emptyset) - H(\{g_1\}) - H(\{g_2\}) + H(\{g_1, g_2\}) \tag{18}$$

- Estimating the effect of variants reduced to observing heights of people with different genotypes.
- **The Möbius inversion links macro observables to micro interactions**
- Let's start decomposing some things!

Part 3: Decompositions of Complex Systems

Goal: convince you that this is *everywhere:*

- Statistics & information theory
- Biology
- Physics
- Game theory
- AI
- *etc...*



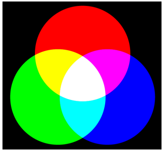
Name of the game: choose a decomposition and a Q, then calculate μ to estimate q.

Decomposing Colours

- Colours are like sets ordered by inclusion.
- $\mu(c, d) = \pm 1$
- If ordering is additive:

$$I_{Magenta} = I_{Red,Blue} = \sum_{c \leq Magenta} \mu(c, Magenta)\ c$$

$$= Magenta - Red - Blue + Black$$

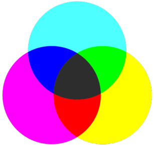


Additive colour mixing

- If ordering is subtractive:

$$I_{Magenta} = \sum_{c \leq Magenta} \mu(c, Magenta)\ c$$

$$= Magenta - White$$



Subtractive colour mixing

Information Theory

- Entropy of a set of random variables $S = \{S_1, S_2, \ldots, S_n\}$:

$$H(S) = -\sum_{s \in \mathcal{S}} p(S = s) \log p(S = s) \tag{19}$$

- **Assume:** the information content of a set of variables decomposes into contributions from subsets. Then

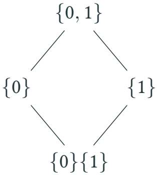
$$H(S) = \sum_{t \subseteq S} I(t) \tag{20}$$

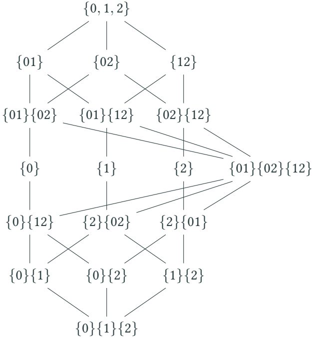
$$\iff \quad I(S) = \sum_{t \leq S} \mu_{\mathcal{P}}(t, S) H(t) = \sum_{t \leq S} (-1)^{|t| - |S|} H(t) \tag{21}$$

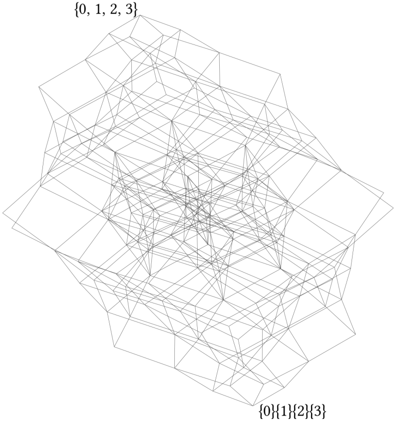
- $I(X, Y) = -H(X) - H(Y) + H(X, Y)$
- $I(X, Y, Z) = H(X) + H(Y) + H(Z) - H(X, Y) - H(X, Z) - H(Y, Z) + H(X, Y, Z)$
- Mutual information is the Möbius inverse of entropy!

Redundancy ordering

- Can we decompose mutual information further?
- Decompose $I(S = \{S_1, \ldots S_n\}, T)$ into *antichains* $\mathcal{A}(S)$ of $\mathcal{P}(S)$. e.g. $\{a, b\}$ and $\{b, c, d\}$.
- For A, B antichains, let $A \leq B$ if for every $b \in B$ there is an $a \in A$ such that $a \subseteq b$.







Partial Information Decomposition

- Information decomposition: $I(S, T) = \sum_{A \in \mathcal{A}(S)} \Pi(A, T)$
- Two source variables:

$$I(\{X_1, X_2\}; Y) = \Pi(\{X_1\}; Y) + \Pi(\{X_2\}; Y) + \Pi(\{X_1\}\{X_2\}; Y) + \Pi(\{X_1, X_2\}; Y) \tag{22}$$

- Decomposes information into *unique*, *redundant*, and *synergistic* contributions.
- Knowledge of $\mu_{\mathcal{A}(S)}$ allows for estimation of each type of information!

$$\Pi(S, T) = \sum_{A \in \mathcal{A}(S)} \mu_{\mathcal{A}(S)}(A, S) I(A, T) \tag{23}$$

- (Upcoming work with Fernando Rosas)
- Commonly used in neuroscience, IIT, etc.

Biology

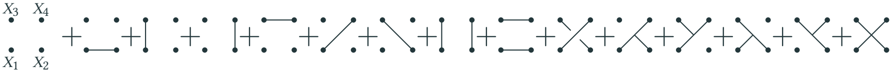
- Say a phenotype F depends on presence of genetic variants $g \subseteq G$:

$$F(g = 1, G \setminus g = 0) = \sum_{s \in \mathcal{P}(g)} I(s) \tag{24}$$

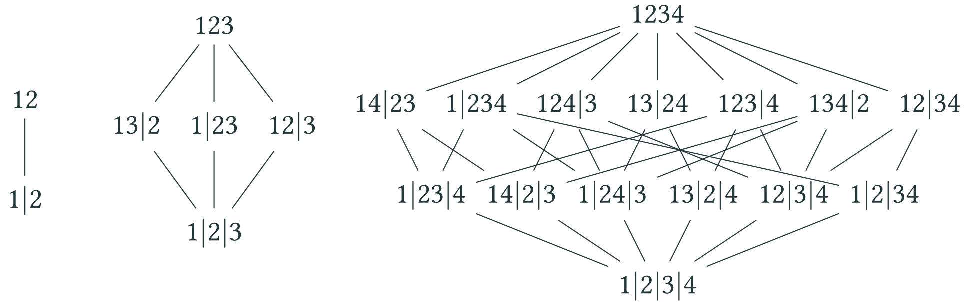
$$\implies I(g_1, g_2, g_3) = F_{111} - F_{110} - F_{101} - F_{011} + F_{100} + F_{010} + F_{001} - F_{000} \tag{25}$$

- **Epistasis:** A measure of how genetic variants interact to produce a phenotype.
- *cf. Sturmfels, Pachter, Beerenwinkel (2007)*: algebraic *vs.* geometric
- **Alternatives:**
  - Gene expression instead of variants $\implies$ Transcript interactions (*High order expression dependencies finely resolve cryptic states and subtypes in single cell data* - AJ et al. 2023)
  - Treatments instead of variants $\implies$ Average treatment effects, drug interactions, etc.

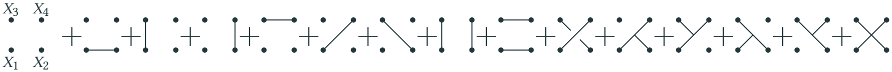
- **Statistical mechanics:** Decompose correlations into physical processes

$\langle X_1 X_2 X_3 X_4 \rangle =$ 

- This is essentially a sum over partitions $\langle X \rangle = \sum_{\pi \in \Pi(X)} \prod_{\pi_i \in \pi} u(\pi_i)$



- **Statistical mechanics:** Decompose correlations into physical processes

$\langle X_1 X_2 X_3 X_4 \rangle =$ 

- This is essentially a sum over partitions $\langle X \rangle = \sum_{\pi \in \Pi(X)} \prod_{\pi_i \in \pi} u(\pi_i)$
- The Möbius function of partitions ordered by refinement is given by

$$\mu_{\Pi(S)}(x, \hat{1}) = (-1)^{|x|-1}(|x|-1)! \tag{26}$$

such that

$$u(X_1, X_2) = \langle X_1 X_2 \rangle - \langle X_1 \rangle \langle X_2 \rangle \tag{27}$$

$$u(X_1, X_2, X_3) = \langle X_1 X_2 X_3 \rangle - \langle X_1 X_2 \rangle \langle X_3 \rangle - \langle X_1 X_3 \rangle \langle X_2 \rangle - \tag{28}$$

$$\langle X_2 X_3 \rangle \langle X_1 \rangle + 2 \langle X_1 \rangle \langle X_2 \rangle \langle X_3 \rangle \tag{29}$$

- **Ursell functions / Scattering amplitudes**

Game Theory

- Players can form coalitions to increase their payoff.
- Is there synergy in coalitions? How much should a player be rewarded for cooperating?
- Value v of a grand coalition S can be decomposed into sub-coalition synergies w:

$$v(S) = \sum_{R \subseteq S} w(R) \tag{30}$$

$$w(R) = \sum_{R \subseteq S} (-1)^{|S|-|R|} v(S) \tag{31}$$

- Shapley: A player should be awarded the average of their contributions to all their coalitions

$$\phi_i = \sum_{R \subseteq S : i \in R} \frac{w(R)}{|R|} \tag{32}$$

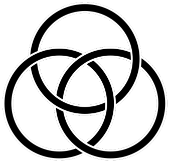
- **Shapley value** – Nobel prize in economics 2012

Summary

Field of Study	Macro Quantity	Decomposition	Micro Quantity/Interactions
Statistics	Moments	Powerset	Central moments
	Moments	Partitions	Cumulants
	Free moments	Non-crossing partitions	Free cumulants
	Path signature moments	Ordered partitions	Path signature cumulants
Information Theory	Entropy	Powerset	Mutual information
	Surprisal	Powerset	Pointwise mutual information
	Joint Surprisal	Powerset	Conditional interactions
	Mutual Information	Antichains	Synergy/redundancy atoms
Biology	Pheno- & Genotype	Powerset	Epistasis
	Gene expression profile	Powerset	Genetic interactions
	Population statistics	Powerset	Synergistic treatment effects
Physics	Ensemble energies	Powerset	Ising interactions
	Correlation functions	Partitions	Ursell functions
	Quantum corr. functions	Partitions	Scattering amplitudes
Chemistry	Molecular property	Subgraphs	Fragment contributions
	Molecular property	Reaction poset	Cluster contributions
Game Theory	Coalition value	Powerset	Coalition synergy
	Shapley value	Powerset	Normalised coalition synergy
Artificial intelligence	Generative model probabilities	Powerset	Feature interactions
	Predictive model predictions	Powerset	Feature contributions
	Dempster-Shafer Belief	Distributive	Evidence weight

Conclusion

- Plato was right: Carve nature at its joints.
- Higher-order interactions inherit their justification from the decomposition.
- Many decompositions allow for a **partial order**.
- *Higher-order* means higher in this partial order!
- The Möbius function of the decomposition solves the inverse problem.
- This is a general construction that appears in many fields.
- **Please let me know if you come across any other examples!**
- Future work: generalised decompositions, categorification, novel interactions.
- Thank you!



References

- Rota, GC *On the foundations of combinatorial theory I. Theory of Möbius functions*
- AJ, 2024 - *A Compositional Approach to Higher-Order Structure in Complex Systems: Carving Nature at its Joints*
- AJ, 2023 - *Higher-Order Interactions and their Duals Reveal Synergy and Logical Dependence Beyond Shannon-Information*
- Beentjes, SV& Khamseh, A, 2020 - *Higher-order interactions in statistical physics and machine learning: A model-independent solution to the inverse problem at equilibrium.*
- AJ, et al. 2023 - *High order expression dependencies finely resolve cryptic states and subtypes in single cell data*
- Beerenwinkel, N & Pachter, L & Sturmfels, B, 2007 - *Epistasis and shapes of fitness landscapes*