Compositional statistical mechanics, entropy and variational free energy (Extended abstract)

Grégoire Sergeant-Perthuis

Compositional statistical mechanics, entropy and variational free energy (Extended abstract)

Grégoire Sergeant-Perthuis
LCQB Sorbonne Université
gregoireserper@gmail.com

Statistical physics is a framework that focuses on the probabilistic description of complex systems [1] (statistical systems). It serves as a rich framework for probabilistic modeling [2–6]. Rigorous Statistical Mechanics is centered on the mathematical study of statistical systems. Central concepts in this field have a natural expression in terms of partially ordered set (poset) shaped diagrams in a category that couples measurable maps and Markov kernels [7]. Such generalization is motivated by the desire to 'compose' in a controlled and computable manner nontrivial statistical systems, and their phases, from simpler ones. Our work follows the line of works that propose new foundations, based on topology and geometry, for probability theory, information theory, and deep learning [8–13], and applies compositional reasoning to engineering [9, 12, 14]. Our contribution is a summary of various results related to a compositional/categorical approach to rigorous Statistical Mechanics [7, 15–21], based on the content of [22, 23].

We showed that statistical systems are particular representations of posets, which we call $\mathscr{A}$-specifications, and expressed their phases, i.e. Gibbs measures, as invariants of these representations. Let us denote **Mes** as the category of measurable spaces with measurable maps as morphisms and **Kern** as the category of measurable spaces with Markov kernels as morphisms. Recall that a Markov kernel $F : E \to E_1$ is a measurable map from the measurable space E to $\mathbb{P}(E_1)$, the measurable space of probability measures over E_1. For G a presheaf with source a poset, we will denote $G(b \leq a)$ as G^a_b.

Definition 1 ($\mathscr{A}$-Specifications)**.** Let $\mathscr{A}$ be a poset, an $\mathscr{A}$-specification is a couple (G,F) of a presheaf and a functor where $G : \mathscr{A}^{op} \to \mathbf{Mes}$ and $F : \mathscr{A} \to \mathbf{Kern}$ are such that for any $a,b \in \mathscr{A}$ with $b \leq a$, $G^a_b \circ F^b_a = \mathrm{id}$.

Definition 2 (Gibbs measures for $\mathscr{A}$-specifications)**.** Let $\gamma = (G,F)$ be an $\mathscr{A}$-specification, we call the Gibbs measures of γ the sections of F: $\mathscr{G}_g(\gamma) := [*,F]_{K,\mathscr{A}}$ where $[*,F]_{K,\mathscr{A}} := \{(p_a \in \mathbb{P}(F(a)), a \in \mathscr{A}) | \quad \forall b \leq a, F^b_a \circ p_b = p_a\}$.

Two central results of rigorous Statistical Mechanics are, firstly, the characterization of extreme Gibbs measure as it relates to the zero–one law for extreme Gibbs measures, and, secondly, their variational principle which states that for translation invariant Hamiltonians, Gibbs measures are the minima of the Gibbs free energy. We showed in [21] how the characterization of extreme Gibbs measures extends to $\mathscr{A}$-specifications. Recent results in categorical probability theory give a characterization of the zero–one law for independent random variables and for Markov chains in a categorical formulation [24, 25]. The zero–one law for extreme Gibbs measures is known to extend the ones of independent random variables and Markov chains [26], so it would be expected that the categorical formulation of extreme Gibbs measures we proposed may also relate to the categorical formulation developed in the cases of independent random variables and Markov chains.

We proposed in [27] an Entropy functional for $\mathscr{A}$-specifications and gave a message-passing algorithm which fix points are critical points of an associated free energy. This algorithm generalized the belief propagation algorithm of graphical models.

Submitted to:
ACT 2024

1 Characterization of extreme Gibbs measures of $\mathscr{A}$-specifications

Consider an $\mathscr{A}$-specification (G,F) and assume the measurable sets $G(a), a \in \mathscr{A}$ are finite. We will say that $F>0$ when for any $a,b \in \mathscr{A}$, such that $b \leq a$, $F(\omega_a|\omega_b)>0$ for any $\omega_b \in G(b), \omega_a \in G(a)$ such that $G^a_b(\omega_a)=\omega_b$; $G\circ F = \mathrm{id}$ requires that $F(\omega_a|\omega_b)=0$ when $G^a_b(\omega_a)\neq \omega_b$. We propose that one candidate that plays the role of the tail σ-algebra for a given specification $\gamma=(G,F)$ is $\lim \sigma(G)$ defined as,

$$\lim \sigma(G) := \{(A_a \in \sigma(G(a)), a \in \mathscr{A}) | \forall a,b \in \mathscr{A}, \quad A_a = G^{a-1}_b(A_b)\} \tag{1}$$

Theorem 1 (Extreme measure characterisation)**.** *Let $\gamma=(G,F)$ be a specification, let $G(a)$ be finite sets for any $a\in\mathscr{A}$, let $F>0$. $\mathscr{G}_g(\gamma)$ is a convex set. Each $\mu \in \mathscr{G}_g(\gamma)$ is uniquely determined by it's restriction to* $\lim\sigma(G)$*. Furthermore μ is extreme in $\mathscr{G}(\gamma)$ if and only if for any $A \in \lim\sigma(G)$, $\forall a \in \mathscr{A}$, $\mu_a(A_a)=0$ or* 1.

2 Entropy of $\mathscr{A}$-specifications and variational free energy

Let $\mathscr{A}$ be a finite poset and $\gamma=(G,F)$ be a specification with $G(a)$ a finite set for any $a\in\mathscr{A}$. We propose the entropy of $Q \in \mathscr{G}_g(\gamma)$ to be $S_{GB}(Q)=\sum_{a\in\mathscr{A}} c(a)S(Q_a)$ with $c(a)$ that relates to the Möbius function of the poset $\mathscr{A}$ which we will introduce just after; $S(Q_a) = -\sum_{\omega_a} Q_a(\omega_a)\ln Q_a(\omega_a)$ is the entropy of Q_a. The variational free energy of a $\mathscr{A}$-specification is defined as $F_{\mathrm{Bethe}}(Q)=\sum_{a\in\mathscr{A}} c(a)\left(\mathbb{E}_{Q_a}[H_a]-S(Q_a)\right)$ with $H_a : G(a)\to\mathbb{R}$ a measurable map. This expression of entropy and free energy is motivated by the Bethe free energy of graphical models and factor graphs, which is an approximation of the Gibbs free energy [28] and is used for (variational) inference on graphical models, factor graphs, etc.

Problem to solve: The optimization problem we want to solve is the following: $\inf_{Q\in\mathscr{G}_g(\gamma)} F_{\mathrm{Bethe}}(Q)$

For $\mathscr{A}$ a finite poset, we call the 'zeta-operator' of $\mathscr{A}$, denoted ζ, the operator from $\bigoplus_{a\in\mathscr{A}}\mathbb{R}$ to $\bigoplus_{a\in\mathscr{A}}\mathbb{R}$ defined as, for any $\lambda \in \bigoplus_{a\in\mathscr{A}}\mathbb{R}$ and any $a\in\mathscr{A}$, $\zeta(\lambda)(a)=\sum\limits_{b\leq a}\lambda_b$. ζ is invertible [29], we denote μ its inverse and its matrix expression $(\mu(a,b), b\leq a)$ defines the Möbius function of $\mathscr{A}$. For a functor G from $\mathscr{A}$ to $\mathbb{R}$-vector spaces, we define μ_G as, for any $a\in\mathscr{A}$ and $v\in\bigoplus_{a\in\mathscr{A}}G(a)$,

$$\mu_G(v)(a) = \sum_{b\leq a}\mu(a,b)G^b_a(v_b) \tag{2}$$

Let $\tilde{G}$ be the presheaf from $\mathscr{A}$ to the category of finite vector spaces defined by for $b\leq a$, $\tilde{G}^a_b : \mathbb{P}(G(a))\to\mathbb{P}(G(b))$ such that $\tilde{G}^a_b(p_a)=G^a_b\circ p_a$ for $p\in\mathbb{P}(G(a))$. We denote G^* the functor obtained by dualizing the morphisms $\tilde{G}^a_b$. Let $FE : \prod_{a\in\mathscr{A}}\mathbb{P}(E_a)\to\prod_{a\in\mathscr{A}}\mathbb{R}$ be such that $FE(Q)=(\mathbb{E}_{Q_a}[H_a]-S(Q_a), a\in\mathscr{A})$; FE sends a collection of probability measures over $\mathscr{A}$ to their Gibbs free energies. For any $Q\in\prod_{a\in\mathscr{A}}\mathbb{P}(E_a)$, let us denote $\mathrm{d}_Q FE$ as the differential of FE at the point Q.

Theorem 2. *Let $\mathscr{A}$ be a finite poset, let $\gamma=(G,F)$ be a $\mathscr{A}$-specification such that $G(a)$ is a finite set for any $a\in\mathscr{A}$. Let $H_a : G(a)\to\mathbb{R}$ be a collection of (measurable) Hamiltonians. The critical points of F_{Bethe} are the $Q\in[*,F]_{K,\mathscr{A}}$ such that,*

$$\mu_{G^*} d_Q FE|_{[*,F]_{K,\mathscr{A}}} = 0 \tag{3}$$

We propose a message passing algorithm to find the critical points of F_{Bethe} for $\mathscr{A}$-specifications; it extends the (General) Belief Propagation in the case of $\mathscr{A}$-specifications (see [22,23,27])

References

[1] D. Ruelle, *Statistical Mechanics*. Imperial College Press, 1999.

[2] M. Mezard and A. Montanari, *Information, Physics, and Computation*. USA: Oxford University Press, Inc., 2009.

[3] T. Lelièvre, M. Rousset, and G. Stoltz, *Free Energy Computations*. Imperial College Press, 2010.

[4] C. Chipot and A. Pohorille, eds., *Free energy calculations: theory and applications in chemistry and biology*. No. 86 in Springer series in chemical physics, Berlin ; New York: Springer, 2007. OCLC: ocm79447449.

[5] M. Wiering and M. van Otterlo, eds., *Reinforcement Learning: State-of-the-Art*, vol. 12 of *Adaptation, Learning, and Optimization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[6] Y. Lecun, S. Chopra, R. Hadsell, M. Ranzato, and F. Huang, *A tutorial on energy-based learning*. MIT Press, 2006.

[7] G. Sergeant-Perthuis, "A categorical approach to statistical mechanics," in *Geometric Science of Information* (F. Nielsen and F. Barbaresco, eds.), (Cham), pp. 258–267, Springer Nature Switzerland, 2023.

[8] M. Gromov, "In a search for a structure, part 1: On entropy.," 2013.

[9] T. Fritz, "A synthetic approach to markov kernels, conditional independence and theorems on sufficient statistics," *Advances in Mathematics*, 2020.

[10] P. Baudot and D. Bennequin, "The homological nature of entropy," *Entropy*, vol. 17, no. 5, pp. 3253–3318, 2015.

[11] J. P. Vigneaux, "Information structures and their cohomology," *Theory and Applications of Categories*, vol. 35, no. 38, pp. 1476–1529, 2020.

[12] T. Fritz and D. I. Spivak, "Internal probability valuations." Workshop Categorical Probability and Statistics, 2020.

[13] J.-C. Belfiore and D. Bennequin, "Topos and stacks of deep neural networks," *ArXiv*, 2021.

[14] T. Fritz and P. Perrone, "A probability monad as the colimit of spaces of finite samples," *Theory and Applications of Categories*, 2019.

[15] G. Sergeant-Perthuis, "Bayesian/graphoid intersection property for factorisation spaces," 2021. `https://arxiv.org/abs/1903.06026`.

[16] G. Sergeant-Perthuis, "Intersection property and interaction decomposition," Apr. 2019. `https://arxiv.org/abs/1904.09017`.

[17] G. Sergeant-Perthuis, "Interaction decomposition for presheaves," Aug. 2020. `https://arxiv.org/abs/2008.09029`.

[18] D. Bennequin, O. Peltre, G. Sergeant-Perthuis, and J. P. Vigneaux, "Extra-fine sheaves and interaction decompositions," Sept. 2020. `https://arxiv.org/abs/2009.12646`.

[19] G. Sergeant-Perthuis, "Interaction decomposition for Hilbert spaces," 2021. `https://arxiv.org/abs/2107.06444`.

[20] G. Sergeant-Perthuis, *Intersection property, interaction decomposition, regionalized optimization and applications.* PhD thesis, Université de Paris, 2021. Link.

[21] G. Sergeant-Perthuis, "Characterization of extreme Gibbs measures for a Categorical Approach to Statistical Mechanics," Feb. 2024. `https://hal.sorbonne-universite.fr/hal-04456412`.

[22] G. Sergeant-Perthuis, "Compositional statistical mechanics, entropy and variational inference," in *Twelfth Symposium on Compositional Structures*, 2024.

[23] G. Sergeant-Perthuis, "Compositional statistical mechanics, entropy and variational inference," 2024. `https://arxiv.org/abs/2403.16104`.

[24] T. Fritz, T. Gonda, and P. Perrone, "De finetti's theorem in categorical probability," *arXiv preprint arXiv:2105.02639*, 2021.

[25] S. Moss and P. Perrone, “A category-theoretic proof of the ergodic decomposition theorem,” *Ergodic Theory and Dynamical Systems*, vol. 43, no. 12, p. 4166–4192, 2023.

[26] H.-O. Georgii, *Gibbs Measures and Phase Transitions*. Berlin, New York: De Gruyter, 2011.

[27] G. Sergeant-Perthuis, “Regionalized optimization,” 2022. `https://arxiv.org/abs/2201.11876`.

[28] J. Yedidia, W. Freeman, and Y. Weiss, “Constructing Free-Energy Approximations and Generalized Belief Propagation Algorithms,” *IEEE Transactions on Information Theory*, vol. 51, pp. 2282–2312, July 2005.

[29] G.-C. Rota, “On the foundations of combinatorial theory I. Theory of Möbius functions,” *Probability theory and related fields*, vol. 2, no. 4, pp. 340–368, 1964.