ANNALES SCIENTIFIQUES
DE L'UNIVERSITÉ DE CLERMONT-FERRAND 2
Série Probabilités et applications

MARC PIRLOT

Variational principle, conjugate convex functions and "the equivalence of ensembles"

Annales scientifiques de l'Université de Clermont-Ferrand 2, tome 87, série *Probabilités et applications*, n° 4 (1985), p. 57-68

<http://www.numdam.org/item?id=ASCFPA_1985__87_4_57_0>

VARIATIONAL PRINCIPLE, CONJUGATE CONVEX FUNCTIONS AND "THE EQUIVALENCE OF ENSEMBLES"

Marc PIRLOT

Summary

In classical statistical mechanics on a lattice, we use the theory of conjugate convex functions to prove a (conjugate) variational principle for the entropy and to give a quite general explanation of the "equivalence of ensembles".

1. Introduction and notations

In this paper we show how the abstract theory of conjugate convex functions (generalizing the Legendre transformation) can be used to formulate and analyse completely the problem of "equivalence of ensembles" in Statistical Mechanics on a lattice.

We show that the results obtained by Lanford [6] are not subordinated to the possibility of constructing a thermodynamic potential (e.g. Helmholtz' free energy) as a thermodynamic limit of local quantities but that they are of a very general nature. Our approach is close to the physicists'one.

Moreover, the results established here for compact spins on Z^{ν} extend without difficulty to any model for which a variational principle is available.

The notion of conjugate convex functions was already used or mentioned in [7], [9] and by Wightman in the preface of [5] but results like our theorems 2 and 3 were not derived there.

In § 2, we remark that the usual variational principle exare conjugate convex functions. From that, follows immediately the "conjugate" variational principle for the entropy (Ruelle and Israël gave much more complicated proofs : see [5]).

The third paragraph provides a key for understanding the equivalence of the ensembles. It is proven under a mild condition, that the states giving a fixed mean energy to some potentials and the value of the free energy (or more general thermodynamic potentials) corresponding to that mean energy, are equilibrium states (i.e. tangent to the pressure). Mathematically, one shows that a "partial" conjugate of the pressure on a subspace of the interactions space is related to the entropy in the expected way.

Finally, we only consider partial conjugates of the pressure on a finite dimensional subspace on which the pressure is strictly convex. We investigate the mild condition mentioned above, draw some physically relevant consequences about the shape of the thermodynamic potentials, give a precise formulation of the equivalence of the ensembles and mention a version of Gibbs Phase rule.

We consider the general model of compact spins on Z^ν. Let Ω_0 be a compact metric space, $\mathfrak{F}_0$, the borelian σ-algebra on Ω_0 and λ_0 a probability on $(\Omega_0, \mathfrak{F}_0)$.

The space of configurations $\Omega = (\Omega_0)^{Z^\nu}$ with the product topology, is compact and metrisable ; the σ-algebra $\mathfrak{F}$ is the product of the σ-algebras $\mathfrak{F}_0$ on each copy of Ω_0.
For Λ a finite part of Z^ν, $\Omega(\Lambda) := (\Omega_0)^\Lambda$, $\mathfrak{F}(\Lambda)$ is the product σ-algebra on $\Omega(\Lambda)$ and λ_Λ the product measure on $(\Omega(\Lambda), \mathfrak{F}(\Lambda))$. We denote by ω_Λ, the restriction of a configuration $\omega \in \Omega$ to Λ.

An interaction potential φ is a family of continuous functions $\varphi_X : \Omega(X) \to \mathbb{R}$, indexed by the finite parts of Z^ν. $\mathcal{B}$ is the Banach space of translation invariant potentials with the norm

$$\|\varphi\| := \sum_{X \ni 0,\ X \text{ finite}} \frac{\sup |\varphi_X(\cdot)|}{|X|}$$

Another space of interest is $\tilde{\mathcal{B}}$, the space of invariant potentials normed by

$$\|\varphi\| := \sum_{X \ni 0,\ X \text{ finite}} \sup |\varphi_X(\cdot)|$$

One defines the usual thermodynamic functions as follows, the limits being taken along sequences of finite boxes Λ tending to infinity in the sense of Van Hove (see [5] for proofs). For $\mu \in \mathcal{E}^I$, the set of translation invariant probabilities on $(\Omega, \mathcal{F})$, one defines entropy $s(\mu)$ by :

$$s(\mu) = \begin{cases} \lim\limits_{\Lambda \to \infty} \dfrac{-1}{|\Lambda|} \displaystyle\int \ln \frac{d\mu_\Lambda}{d\lambda_\Lambda}(\omega_\Lambda)\, \mu_\Lambda(d\omega_\Lambda) \\ \qquad \text{if this expression makes sense} \\ -\infty \text{ otherwise} \end{cases}$$

Entropy is an affine u.s.c. function ; $-\infty \leqslant s(\mu) \leqslant 0$.

If $\mu \in \mathcal{E}^I$ and $\varphi \in \mathcal{B}$, the mean energy $e(\mu, \varphi)$ is the affine continuous function on $\mathcal{B}$ defined by :

$$e(\mu, \varphi) = \int \sum_{X \ni 0,\ X \text{ finite}} \frac{\varphi_X(\omega_X)}{|X|} \mu(d\omega)$$

The pressure P is a continuous (Lipschitzian) convex function defined for any φ in $\mathcal{B}$ by :

$$P(\varphi) = \lim_{\Lambda\to\infty} \frac{1}{|\Lambda|} \ln \int \exp [- V_\Lambda^\varphi (\omega_\Lambda)] \lambda_\Lambda (d\, \omega_\Lambda)$$

with $V_\Lambda^\varphi (\omega_\Lambda) = \sum_{X\subset\Lambda} \varphi_X (\omega_X)$

THEOREM 1. (Variational principle, [5]). For any φ in $\mathcal{B}$,

$$\sup [s(\mu) - e(\mu, \varphi) \; ; \mu \in \mathcal{E}^I] = P(\varphi)$$

The supremum is reached by all equilibrium states for φ.

2. Conjugate variational principle

Let $\mathcal{B}$ be a Banach space (in fact a locally convex linear topological space is enough) and $\mathcal{B}^*$ its dual. Let $\mathcal{B}^*$ be endowed the weak*-topology and $\mathcal{B}$ with the weak topology. With those topologies, $\mathcal{B}$ and $\mathcal{B}^*$ are weak duals of each other (see e.g. [11], theor. 1, p. 112). There is a one to one correspondence between the set $\Gamma(\mathcal{B})$ of weakly closed (l.s.c.), proper (not identically equal to $+\infty$), convex functions on $\mathcal{B}$ and the set $\Gamma(\mathcal{B}^*)$ of weakly* closed, proper, convex functions on $\mathcal{B}^*$. Let f belong to $\Gamma(\mathcal{B})$; one defines a function f^* on $\mathcal{B}^*$ by :

$$f^*(x^*) = \sup [< x,x^* > - f(x), \; x \in \mathcal{B}] \qquad (1)$$

f^* is called the conjugate of f. As is well-known ([3], [8]), f^* belongs to $\Gamma(\mathcal{B}^*)$ and

$$f(x) = \sup [< x,x^* > - f^*(x^*), \; x^* \in \mathcal{B}^*] \qquad (2)$$

Reciprocally, starting with g of $\Gamma(\mathcal{B}^*)$, we define

$$g^*(x) = \sup [< x,x^* > - g(x^*), \; x^* \in \mathcal{B}^*]$$

and due to the weak duality between $\mathcal{B}$ and $\mathcal{B}^*$, it is possible to prove, in the same way as for f^*, that g^* belongs to $\Gamma(\mathcal{B})$ and

$$g(x^*) = \sup [< x, x^* > - g^*(x), \; x \in \mathcal{B}]$$

Let us define the weak* lower semi continuous, proper, convex function t (α) on $\mathcal{B}^*$, as

$$t\,(\alpha) = \begin{cases} - s(\mu) \text{ if there is } \mu \in \mathcal{E}^I : - e(\mu, \varphi) = \alpha(\varphi), \; (\varphi \in \mathcal{B}) \\ + \infty \quad \text{otherwise} \end{cases}$$

The definition is unambiguous as two different μ cannot yield the same α. By the direct variational principle (theor. 1) :

$$P(\varphi) = \sup [\alpha(\varphi) - t\,(\alpha) \; ; \; \alpha \in \mathcal{B}^*]$$

As P is the conjugate of t, P is in $\Gamma(\mathcal{B})$ and

$$t\,(\alpha) = \sup [\alpha(\varphi) - P(\varphi), \; \varphi \in \mathcal{B}]$$

One sees that t (α) is finite iff α is P-bounded, i.e. if there exists a constant c in $\mathbb{R}$ such that for any φ in $\mathcal{B}$:

$$P(\varphi) \geqslant \alpha(\varphi) - c$$

The following three properties are all immediate consequences of what precedes. Corollary 2.1 is theorem II.3.4 in [5] and corollary 2.2 implies theorem II.1.2 in [5].

<u>THEOREM 2</u>. <u>The conjugate of P is the function t : $\mathcal{B}^* \to \mathbb{R}$</u> :

$$t\,(\alpha) = \begin{cases} - s(\mu) \text{ if there is } \mu \in \mathcal{E}^I : - e(\mu, \varphi) = \alpha(\varphi), \; (\varphi \in \mathcal{B}) \\ + \infty \quad \text{otherwise} \end{cases}$$

<u>For any θ in $\mathcal{B}$, the conjugate of $P(\theta + .)$ is $t(\alpha) - \alpha(\theta)$.</u>

<u>COROLLARY 2.1</u> (<u>Conjugate variational principle</u>) <u>For every μ in $\mathcal{E}^I$</u> :

$$s(\mu) = \inf [P(\varphi) + e(\mu, \varphi) \; ; \; \varphi \in \mathcal{B}]$$

<u>COROLLARY 2.2</u> <u>$\alpha \in \mathcal{B}^*$ is P-bounded iff there exists μ in $\mathcal{E}^I$ such that $\alpha(\varphi) = - e(\mu, \varphi)$ for all φ in $\mathcal{B}$ and $s(\mu)$ is finite iff $t(\alpha)$ is finite</u>.

3. <u>Partial conjugates of the pressure</u>

Let $\mathcal{B}'$ be a linear subspace of $\mathcal{B}$. For instance $\mathcal{B}'$ is finite dimensional and generated by the independent components of $\overline{\psi} = (\psi_1, \ldots, \psi_n)$, ψ_i in $\mathcal{B}$. Physically relevant situations in the model where $\Omega_0 = \{0,1\}$ are the :

a) canonical description : $n = 1$, $\psi_1(\omega_X) = \begin{cases} 1 \text{ if } |X| = 1 \text{ and } \omega_X = 1 \\ 0 \text{ otherwise} \end{cases}$

b) microcanonical description : $n = 2$, ψ_1 is the same as in the canonical description and ψ_2 describes the interactions between the particles.

The partial conjugate of P on $\theta + \mathcal{B}'$, with θ in $\mathcal{B}$, is the convex, l.s.c. function $F(\theta, \alpha')$, $\alpha' \in \mathcal{B}'^*$, defined by :

$$F(\theta, \alpha') = \sup [\alpha'(\varphi) - P(\theta + \varphi) \; ; \; \varphi \in \mathcal{B}']$$

The supremum defining F is not always attained (see theor. 4 below, for the case dim $\mathcal{B}'$ is finite).

THEOREM 3. $F(\theta, \alpha') = \inf [t(\alpha) - \alpha(\theta) ; \alpha \in \mathcal{B}^*, \alpha|\mathcal{B}' = \alpha']$ and the infimum is reached for every α' in $\mathcal{B}'^*$.

Proof. It is clear (by theor. 1) that $F(\theta, \alpha') \leqslant t(\alpha) - \alpha(\theta)$ for any $\alpha \in \mathcal{B}^*$ agreeing with α' on $\mathcal{B}'$. Suppose first that α' is $P(\theta + .)$-bounded on $\mathcal{B}'$, i.e. there exists $c \in \mathbb{R}$ such that $P(\theta + \varphi) \geqslant \alpha'(\varphi) - c$ for every φ in $\mathcal{B}'$. This implies $F(\theta, \alpha')$ is finite for any φ in $\mathcal{B}'$:

$$\alpha'(\varphi) \leqslant P(\theta + \varphi) + F(\theta, \alpha')$$

The second member is convex, continuous in φ so that by Hahn-Banach's theorem, there is $\alpha \in \mathcal{B}^*$ agreeing with α' on $\mathcal{B}'$ and satisfying the same inequality as α' but for any φ in $\mathcal{B}$. So, α is P-bounded with constant $F(\theta, \alpha') + \alpha(\theta)$ and :

$$F(\theta, \alpha') \geqslant \alpha(\psi) - P(\psi) - \alpha(\theta) \qquad (\psi \in \mathcal{B})$$

$$\begin{aligned} F(\theta, \alpha') &\geqslant \sup [\alpha(\psi) - P(\psi) ; \psi \in \ ] - \alpha(\theta) \\ &= t(\alpha) - \alpha(\theta) \end{aligned}$$

This proves the theorem when α' is $P(\theta + .)$-bounded on $\mathcal{B}'$. If it is not, $F(\theta, \alpha')$ is infinite and no $\alpha \in \mathcal{B}^*$ agreeing with α' on $\mathcal{B}'$ is P-bounded so that $t(\alpha)$ is also infinite for those α (theor. 2 and cor. 2.2).

COROLLARY 3.1 If there is $\varphi \in \mathcal{B}'$ such that $F(\theta, \alpha') = \alpha'(\varphi) - P(\theta + \varphi)$, then there exists at least one probability $\mu \in \mathcal{E}^I$ satisfying $- e(\mu, \varphi) = \alpha'(\varphi)$ for all φ in $\mathcal{B}'$ and such that

$$F(\theta, \alpha') = e(\mu, \theta) - s(\mu) = \min [e(\nu, \theta) - s(\nu) ; \nu \in \mathcal{E}^I, - e(\nu, \varphi) = \alpha'(\varphi) \text{ for all } \varphi \text{ in } \mathcal{B}']$$

Any such probability μ is an equilibrium state for the potential $\theta + \varphi$.

From the physical point of view, it is a nuisance that the supremum defining $F(\theta, \alpha')$ is in general attained for several potentials φ of $\mathcal{B}'$. It is due to the fact that P is not strictly convex on $\mathcal{B}$ (nor on $\tilde{\mathcal{B}}$).
From the results stated in [10], one infers that P will be strictly convex on a linear subspace $\mathcal{D}$ of $\tilde{\mathcal{B}}$ (or on $\theta + \mathcal{D}$, $\theta \in \tilde{\mathcal{B}}$) if $\mathcal{D}$ contains at most one potential of each class of physically equivalent potentials (for a definition, see [10]). For instance in the case where $\Omega_0 = \{0,1\}$, P is strictly convex on the space of lattice gas potentials ([4]).

4. Consequences

a) In the sequel, we consider a subspace $\mathcal{D}$ on which P is strictly convex and which is generated by a family $\bar{\psi} = (\psi_1, \ldots, \psi_n)$ of n independent potentials of $\tilde{\mathcal{B}}$. A linear form α on $\mathcal{D}$ is given by a vector $\bar{u}$ of $\mathbb{R}^n$ such that $\bar{u} = -\alpha(\bar{\psi})$. By theorem 3, we have :

$$F(\theta, \bar{u}) := \sup [-\bar{w}.\bar{u} - P(\theta + \bar{w}.\bar{\psi}) \; ; \; \bar{w} \in \mathbb{R}^n]$$
$$= \min [t(\alpha) - \alpha(\theta) \; ; \; \alpha \in \mathcal{B}^*, \alpha(\bar{\psi}) = -\bar{u}]$$

Denote by $K(\bar{\psi})$ the set $\{e(\mu, \bar{\psi}) \in \mathbb{R}^n \; ; \; \mu \in \mathcal{E}^I, s(\mu) > -\infty\}$. This is precisely the set of points $\bar{u}$ of R^n on which $F(\theta, \bar{u})$ is finite (by cor. 2.2) for all θ in $\tilde{\mathcal{B}}$. As P is strictly convex on $\theta + \mathcal{D}$, by [8], theor. 26.3, $F(\theta, .)$ is essentially smooth, that is :

1) int $K(\bar{\psi})$ is not empty ;

2) $F(\theta, .)$ is differentiable throughout int $K(\bar{\psi})$;

3) the norm of grad $F(\theta, \bar{u})$ tends to infinity as $\bar{u}$ tends to the boundary of $K(\bar{\psi})$.

The gradient of $F(\theta, \overline{u})$ is the vector $\overline{v}$ of R^n such that

$$P(\theta + \overline{v}.\overline{\psi}) + \overline{v}.\overline{u} = \inf [P(\theta + \overline{w}.\overline{\psi}) + \overline{w}.\overline{u} \ ; \ w \in \mathbb{R}^n]$$

THEOREM 4. The supremum defining $F(\theta, \overline{u})$ is exactly attained on the interior of the convex bounded set $K(\overline{\psi})$.

Proof. In view of the remarks made before, it only remains to prove that if $\overline{u}$ belongs to the boundary of $K(\overline{\psi})$, the supremum may not be attained. Suppose on the contrary that the supremum is reached for some $\overline{v} \in \mathbb{R}^n$. There exists $\overline{\gamma} \in \mathbb{R}^n$ such that $\overline{\gamma}.\overline{u} = \min [\overline{\gamma}.\overline{z} \ ; \ \overline{z} \in K(\overline{\psi})]$; $\overline{\gamma}.\overline{u}$ defines a subgradient of the function $P(k): = P(\theta + \overline{v}.\overline{\psi} + k\ \overline{\gamma}.\overline{\psi}), k \in \mathbb{R}$, in $k = 0$. As $P(k)$ is strictly convex and finite for all k in $\mathbb{R}$, its left and right derivatives $P'_-(k)$ and $P'_+(k)$ are increasing functions. So, there is a point k where $- \overline{\gamma}.\overline{u} < P'_-(k) \leq P'_+(k)$. But $P'_+(k) = - \overline{\gamma}.\overline{z}$ for a certain $\overline{z}$ in $K(\overline{\psi})$, because equilibrium states are invariant probabilities with finite entropy. This is in contradiction with the hypothesis on $\overline{\gamma}$.

COROLLARY 4.1. If μ is a translation invariant Gibbs state, then $e(\mu, \overline{\psi})$ is an interior point of $K(\overline{\psi})$.

Proof. Let μ be an invariant Gibbs state associated to $\varphi \in \mathfrak{B}$; then μ defines a subgradient to P in $\theta + \overline{v}.\overline{\psi}$ for $\theta = \varphi - \overline{v}.\overline{\psi}$ and the supremum defining $F(\theta, e(\mu, \overline{\psi}))$ is attained.

Remark. In the case where Ω_0 is a finite set and λ_0 the normalized counting measure, the set $K(\overline{\psi})$ is compact as $\mathcal{E}^I$ is compact and $s(\mu)$ is a bounded function.

b) Implications for the thermodynamic formalism

According to theorem 3, a function like $F(\theta, \overline{u})$ may be interpreted as a thermodynamic potential (see [1]). For simlicity suppose $n = 1$ so that $\psi = (\psi_1)$. If there is a phase transition in θ and if ψ_1 is an order parameter for that transition, then the graph of $F(\theta, u_1)$ presents a flat (horizontal) part. To compare with the physical usual formalism, see for instance [2] : the extreme values of u_1 on the flat part of $F(\theta, u_1)$ coïncide with the two minima of the function used by the physicists.

In fact, this provides a better description than the usual one, as can be seen with the v_1-u_1 diagrams (e.g. p-V diagrams, ...) relating two conjugate parameters (intensive v_1, extensive u_1) : Maxwell's equal area rule has not to be invoked as the flat part is naturally obtained. The v_1-u_1 diagram is the representation in (u_1, v_1)-axis of the set of couples $(u_1 , - \frac{\partial F}{\partial u_1})$ which are the couples minimizing $F(\theta, u_1) + v_1 u_1$ or, by conjugation, $P(\theta + v_1 \psi_1) + v_1 u_1$.

c) Equivalence of ensembles

A possible interpretation of the "ensembles" and their "equivalence" is the following. Consider that the fundamental problem of statistical mechanics is to determine the couples (φ, μ), $\varphi \in \tilde{\mathcal{B}}$, $\mu \in \mathcal{E}^I$, associated to a given situation. An extreme case is the grand canonical description where φ is completely known (up to temperature, chemical potential, ... : all intensive relevant parameters are known) and μ totally unknown ; the set $\mathcal{G}(\varphi)$ of associated probabilities is the set of Gibbs states for φ (here we only consider the set $\mathcal{G}_0(\varphi)$ of invariant Gibbs States). A generalization of the canonical or microcanonical setting is the case where one only knows that the right potential is a linear combination of some potentials ψ_1, ... ψ_n (generating $\mathcal{D}$) with unknown coefficients (intensive parameters) ; but on the other hand, one knows the values of the extensive parameters conjugated to the unknown intensive parameters, i.e. one knows the mean $\overline{u}$ of the observables associated to $\overline{\psi} = (\psi_1, \ldots, \psi_n)$; the set $\mathcal{C}_0(\theta + \mathcal{D}, \overline{u})$ of (invariant) states associated to the problem is composed of the probabilities $\mu \in \mathcal{E}^I$ satisfying $e(\mu, \overline{\psi}) = \overline{u}$ and $e(\mu, \theta) - s(\mu) = \inf [e(\nu, \theta) - s(\nu) ; \nu \in \mathcal{E}^I, e(\nu, \overline{\psi}) = \overline{u}]$. The other extreme case ("absolute microcanonical") would be a complete knowledge of μ and a total ignorance about φ ; it is what lies under the conjugate variational principle (cor. 2.1).

Corollary 3.1 and theorem 4 allow to analyse the connections between $\mathcal{G}_0(\theta + \overline{v}.\overline{\psi})$ and $\mathcal{C}_0(\theta + \mathcal{D}, \overline{u})$.

1. If $\overline{u}$ is an interior point of $K(\overline{\psi})$, we have in genral $\mathcal{C}_0(\theta + \mathcal{D}, \overline{u}) \subseteq \mathcal{G}_0(\theta + \overline{v}.\overline{\psi})$ for $\overline{v}$ conjugated to $\overline{u}$. Both descriptions are equivalent iff $|\mathcal{G}_0(\theta + \overline{v}.\overline{\psi})| = 1$ or, in the presence of phase transition, if all probabilities of $\mathcal{G}_0(\theta + \overline{v}.\overline{\psi})$ give the same value to $e(., \overline{\psi})$, i.e. if there is no order parameter in $\mathcal{D}$ for the transition.

2. If $\overline{u}$ is not in int $K(\overline{\psi})$, no "grand canonical" description can be given but $\mathcal{C}_0(\theta + \mathcal{D}, \overline{u})$ is non-void when $\overline{u}$ belongs to the boundary of $K(\overline{\psi})$.

d) Gibbs phase rule

One can prove without change but in our much more general setting, the main result of [7], theorem 1 : it says essentially that if $\mathcal{D} = \text{vect}\{\psi_1, \dots \psi_n\}$ contains a representative set of order parameters for a phase transition in θ (i.e. two invariant equilibrium states for θ that don't differ on $\mathcal{D}$ are equal), then the number of pure phases (ergodic equilibrium states) for θ is less than or equal to $n+1$.

REFERENCES

1. BOCCARA, N., Les principes de la thermodynamique classique, Paris, P.U.F., Coll. Sup. 1968.
2. BOCCARA, N., Symétries brisées, Paris, Hermann, 1976.
3. EKELAND, I., TEMAN, R., Analyse convexe et problèmes variationnels, Paris, Dunod-Gauthier-Villars, 1974.
4. GRIFFITHS, R.B., RUELLE, D., Strict convexity ("continuity" of the pressure in lattice systems. Com. math. phys. 23, 169-175 (1971).
5. ISRAEL, R.B., Convexity in the theory of lattice gases, Princeton Series in Physics, Princeton Univ. Press, 1979.
6. LANFORD III, O.E., Entropy and equilibrium states in classical statistical mechanics, in Statistical mechanics and mathematical problems, Ed. A. Lenard, Lect. notes in phys. 20, Springer 1972.
7. MARTIN-LOF, A., The equivalence of ensembles and Gibbs phase rule for classical lattice systems, J. Stat. Phys. 20, 557-571 (1979).
8. ROCKAFELLAR, R.T., Convex analysis, Princeton Univ. Press 1970.
9. ROOS, H., Consequences of convexity : a variational principle for the pressure and related results, Reports math. phys. 14, 147-160 (1978).
10. RUELLE, D., Thermodynamic formalism, Addison-Wesley, 1978
11. YOSIDA, K., Functional analysis, Springer 1966 (2nd printing).

M. PIRLOT
Université de Mons
Place Warocqué, 17, B

7000 MONS
Belgique

Reçu en Mars 1985