

Full paper

# Lagrangian submanifolds generated by the Maximum Entropy principle

Marco Favretti

Dipartimento di Matematica Pura ed Applicata, Università di Padova,  
via Belzoni, 7 – 35131 Padova Italy  
E-mail: favretti@math.unipd.it

Received: 25 October 2004 / Accepted: 12 January 2005 / Published: 12 January 2005

---

**Abstract:** We show that the Maximum Entropy principle (E.T. Jaynes, [8]) has a natural description in terms of Morse Families of a Lagrangian submanifold. This geometric approach becomes useful when dealing with the M.E.P. with nonlinear constraints. Examples are presented using the Ising and Potts models of a ferromagnetic material.

**Keywords:** Symplectic geometry; Maximum Entropy principle; Thermodynamics of mechanical systems; Ising and Potts models

**MSC 2000 codes:** 53D12; 62B10; 82B20

**PACS codes:** 5.20Gg; 5.50.+q; 5.70.Fh

---

### 1 A synopsis of Morse families

Let  $M$  be a smooth even-dimensional manifold and  $\omega$  a non-degenerate closed two-form. The couple  $(M, \omega)$  will be called a *symplectic* manifold. A submanifold  $S \subset M$  such that  $2\dim S = \dim M$  and that  $\omega|_S = 0$  is a *Lagrangian* submanifold of  $M$ .

As an example, given a manifold  $Q$ , its cotangent bundle  $T^*Q$  equipped with the two-form  $\omega = dp \wedge dq$ , where  $(q, p)$  is a local chart of  $T^*Q$ , is a symplectic manifold. A function  $G : Q \rightarrow \mathbb{R}$  defines a Lagrangian submanifold of  $T^*Q$

$$\Lambda_G = \text{graph}(dG) = \{(q, dG(q)) : q \in Q\} \subset T^*Q$$

( $\dim \Lambda_G = \dim Q$  and  $d(dG \wedge dq) = 0$ ) that is *transversal to the fibers* of  $\pi : T^*Q \rightarrow Q$  since  $\pi|_{\Lambda_G}$  is a diffeomorphism.

Let us consider a family of functions over  $Q$  depending on  $k$  parameters

$$G : Q \times U \longrightarrow \mathbb{R}, \quad U \subset \mathbb{R}^k$$

and let us denote with

$$\mathcal{E} = \{(q, u) \in Q \times U : \frac{\partial G}{\partial u} = 0\}$$

the critical set (the set of critical points over the fibers of  $Q \times U \rightarrow Q$ ).

The above introduced family is a *Morse family* (see e.g. Weinstein, [12]) of functions if  $\mathcal{E}$  is a *regular* set in  $Q \times U$ , i.e. iff the following local condition hold on  $\mathcal{E}$

$$\text{rk} \left[ G_{qu} \ G_{uu} \right] \Big|_{(q,u) \in \mathcal{E}} = \max = \dim U = k \tag{1}$$

where  $G_u$  denotes partial derivative of  $G$  with respect to  $u$ .

**Proposition 1.1 (Weinstein, [12] Lect. 6)** *Let  $G$  be a Morse family. Then*

$$\Lambda = \{(q, p) \in T^*Q : p = \frac{\partial G}{\partial q} \quad \text{and } (q, u) \in \mathcal{E} \text{ for some } u \in U\}$$

*is a Lagrangian submanifold<sup>1</sup> of  $T^*Q$ , generated by  $G$ .*

If the above rank condition defining a Morse family is satisfied by the square matrix  $G_{uu}$ , i.e.

$$\text{rk} G_{uu} = \max = k \text{ on } \mathcal{E}, \tag{2}$$

then, by the implicit function theorem, locally at  $q$ ,  $\mathcal{E}$  is the graph of a function  $g_W : W \subset Q \rightarrow U$ ,  $u = g_W(q)$  and setting  $\tilde{G} = G \circ g_W$  one has that

$$\frac{\partial \tilde{G}}{\partial q}(q) = \frac{\partial G}{\partial q}(q, g(q)) + \frac{\partial G}{\partial u} \frac{\partial g}{\partial q}(q, g(q)) = \frac{\partial G}{\partial q}(q, g(q)), \quad \text{for } q \in W. \tag{3}$$

---

<sup>1</sup>An extension of Weinstein' result to the infinite dimensional Banach space setting is contained in [6].

**Proposition 1.2** *If the above condition (2) holds we say that all the parameters  $u$  can be eliminated. Hence, on  $W$ ,  $\Lambda \cap T^*W$  coincides with  $\text{graph}(d\tilde{G})$  therefore  $\Lambda$  is a Lagrangian submanifold locally transversal to the fibers of  $\pi$ .*

If  $G$  is a Morse family, then the set  $\Gamma \subset Q$  of the points  $q \in Q$  such that at  $(q, p) \in \Lambda$  the lagrangian submanifold  $\Lambda$  is not transversal to the fibers of  $\pi$  is called the *caustic* of  $\Lambda$  and it is given by

$$\Gamma = \{q \in Q : \det G_{uu}(q, u) = 0, \text{ and } (q, u) \in \mathcal{E}\}. \tag{4}$$

### 1.1 An example: Lagrange multipliers method

We restate the classical Lagrange multipliers method for the constrained extremization of a differentiable function in terms of Morse families.

Let  $M$  be a  $n$ -dimensional smooth manifold and  $f$  be a smooth real function defined on it. Let  $\zeta : M \rightarrow \mathbb{R}^k$ ,  $k \leq n$  be a smooth submersion i.e. verifying

$$rk \, d\zeta(m) = \max = k \quad \forall m \in M.$$

As it is well known,  $m^* \in M$  is an extremal point,  $df(m^*) = 0$ , of  $f$  on the fiber  $\zeta^{-1}(c)$ ,  $c \in \text{Range}(\zeta) \subset \mathbb{R}^k$  if and only if, upon introducing the smooth map

$$\begin{aligned} G &: \text{Range}(\zeta) \times (\mathbb{R}^k \times M) \rightarrow \mathbb{R}, \\ G(c, \lambda, m) &:= (\zeta(m) - c) \cdot \lambda - f(m) \end{aligned}$$

the couple  $(m^*, \lambda)$  satisfies the system of equations (gradient system)

$$\begin{cases} G_\lambda(c, \lambda, m) = \zeta(m) - c = 0, \\ G_m(c, \lambda, m) = d\zeta^T(m) \cdot \lambda - df(m) = 0. \end{cases} \tag{5}$$

Note that, if the hypothesis  $rk \, d\zeta = \max$  holds, by (5)<sub>2</sub> we can associate to the constrained extremum  $m^*$  a uniquely defined Lagrange multiplier  $\lambda^*$ .

It is straightforward to realize that the set of solutions of the above system of equations (5) when  $c$  varies in  $\text{Range}(\zeta)$  has the structure of a critical set with respect to the projection  $(c, \lambda, m) \mapsto c$ , i.e.

$$\mathcal{E} = \{(c, \lambda, m) \in \text{Range}(\zeta) \times (\mathbb{R}^k \times M) : G_\lambda = 0, G_m = 0\}. \tag{6}$$

Moreover, it is easy to see that the condition that  $G$  be a Morse family reads  $u = (\lambda, m)$

$$rk \left[ G_{cu} \ G_{uu} \right]_{|\mathcal{E}} = rk \left( \begin{array}{ccc} -I_k & 0 & d\zeta \\ 0 & d\zeta^T & G_{mm} \end{array} \right)_{|\mathcal{E}} = \max = k + n$$

where  $I_k$  is the identity matrix in  $\mathbb{R}^k$  and  $G_{mm} \in Sym(n)$  is the Hessian matrix of  $G$  with respect to the  $m$  variables (summation over repeated indices is understood)

$$(G_{mm})_{ij}(\lambda, m) = \lambda_L \frac{\partial^2 \zeta_L(m)}{\partial m_i \partial m_j} - \frac{\partial^2 f(m)}{\partial m_i \partial m_j}. \tag{7}$$

By Proposition 1.1, if  $G$  is a Morse family, the set

$$\Lambda_G = \{(c, \mathbf{p}) \in T^*\mathbb{R}^k : \mathbf{p} = G_c = -\lambda \text{ and } (c, \lambda, m) \in \mathcal{E}\} \tag{8}$$

is a Lagrangian submanifold of  $T^*\mathbb{R}^k$ .

The local condition (2) for the elimination of all the parameters is, in this setting,

$$rk G_{uu}|_{\mathcal{E}} = rk \begin{pmatrix} 0 & d\zeta \\ d\zeta^T & G_{mm} \end{pmatrix}|_{\mathcal{E}} = max = k + n. \tag{9}$$

A sufficient condition (not necessary in the general case) for the elimination of all the parameters is given by the following Proposition (see Bertekas,[3] p. 69)

**Proposition 1.3** *If the symmetric matrix  $G_{mm}(\lambda, m) \in Sym(n)$  in (7) is (positive or negative) definite on  $\ker d\zeta(m) \subset \mathbb{R}^n$  for  $(c, \lambda, m) \in \mathcal{E}$ , then the square matrix  $G_{uu}$  in (9) has maximal rank, and hence all the parameters  $u = (\lambda, m)$  can be eliminated and the Lagrangian submanifold in (8) is, locally at  $c$ , transversal to the fibers of  $\pi$ .*

Proof. We have to show that  $G_{uu}z = 0$  implies that  $z = 0$ , where  $z \in \mathbb{R}^{k+n}$ . Now, setting  $z^T = (w, v) \in \mathbb{R}^{k+n}$ , we have that

$$G_{uu}z = 0 \iff \begin{cases} d\zeta v = 0, \\ d\zeta^T w + G_{mm}v = 0. \end{cases}$$

Now, if  $v = 0$ , the equation  $d\zeta^T w = 0$  implies that  $w = 0$  since  $d\zeta^T$  is injective ( $d\zeta$  is surjective by hypothesis) and we are done. As a reductio ad absurdum hypothesis, suppose that  $v \neq 0$  and  $d\zeta v = 0$ , i.e.  $v \in \ker d\zeta$ . The equation  $d\zeta v = 0$  implies  $v^T d\zeta^T = 0^T$ ; by multiplying the second equation by  $v^T \neq 0$  we get

$$v^T d\zeta^T w + v^T G_{mm}v = v^T G_{mm}v = 0 \quad \text{with } v \neq 0$$

which is an absurdum since  $G_{mm}$  is a (positive on negative) definite matrix on  $\ker d\zeta(m)$ . □

If all the parameters can be eliminated, then, locally at  $c$ , the critical set  $\mathcal{E}$  is the graph of a function  $g(c) = (\tilde{\lambda}(c), \tilde{m}(c))$  and, since (5)<sub>1</sub> holds, locally at  $c$ , the Morse family  $\tilde{G} = G \circ g$  reduces to

$$\tilde{G}(c) = G(c, \tilde{\lambda}(c), \tilde{m}(c)) = (\zeta(\tilde{m}(c)) - c) \cdot \lambda - f(\tilde{m}(c)) = -f(\tilde{m}(c)). \tag{10}$$

Remark. The above sufficient condition involves the "second order" (Hessian) matrix  $G_{mm}$  in (7). In case  $\zeta$  is a linear fibration, the above condition is satisfied if  $Hess(f)$  is a definite matrix on  $\ker d\zeta(m)$ .

Example. In case that  $f$  is the potential energy of a mechanical system subject to conservative forces,  $M$  is the space of configurations of the system, and  $\zeta : M \rightarrow \mathbb{R}^k$  is the fibration defining an ideal (workless) constraint, the above introduced Lagrangian submanifold  $\Lambda$  is the set of couples  $(c, -\lambda)$  respectively value of the constraint and reaction force of the constraint in the equilibrium configuration  $m$ . If for every given  $c$  there is a *unique* equilibrium configuration  $m = \tilde{m}(c)$  then the value of the reaction force  $\lambda = \tilde{m}(c)$  is a function of  $c$ . This situation is an example of the case where all the parameters can be *globally* eliminated.

## 1.2 Global transversality for $\Lambda$

The above condition (2) states that the set of critical points  $(m, \lambda)$  over the fiber  $\zeta^{-1}(c)$  is made of isolated points. If the condition of the Proposition 1.3 above hold, then all the critical points are local maxima (or minima) of  $f$ . We now look at the hypothesis that imply the existence of a *unique* maximum for  $f$  on  $\zeta^{-1}$ , so that the Lagrangian submanifold  $\Lambda$  is *globally* transversal to the fibers of  $\pi$ .

A sufficient condition is the following (see [2], p.136):

**Proposition 1.4** *If the matrix  $G_{uu}$  is positive definite on some bounded convex domain  $D$ , then the map  $G_u$  is globally univalent on  $D$ , hence the equation  $G_u = 0$  has at most one solution in  $D$ .*

## 2 The M. E. P. with linear constraints.

The Maximum Entropy principle (MEP) (Jaynes, [8]) is a general method of statistical inference that it is able to single out a *unique* probability distribution over the space of possible states  $i = 1, \dots, n$  of a system (to be considered as outcomes of  $n$  independent, identically distributed variables) among those compatibles with the results of measurements made on the system, usually in the form of average values of observables defined for the system.

Let  $\chi = \{1, 2, \dots, n\}$  be the (finite) set of possible states (outcomes) of a system. As an example, one can think  $i = (q_i, p_i)$  and  $\chi$  is a finite discretization in cells of the phase space of an Hamiltonian system.

Let  $\varphi_L^i = \varphi_L(i)$  be the value assumed by the *observable*<sup>2</sup>  $\varphi_L : \chi \rightarrow \mathbb{R}$  on the state  $i \in \chi$ ,  $L = 1, \dots, k \leq n$ .

The  $n \times k$  matrix  $\varphi_L^i$  defines a *linear fibration* from  $\mathbb{R}^n$  to  $\mathbb{R}^k$  if the observables are independent, i.e.

$$rk \varphi = max = k. \quad (11)$$

A statistical state over  $\chi$  (an *ensemble* in the Statistical Mechanics language) is a probability vector  $p = (p_1, \dots, p_n) \in \mathbb{R}_+^n$ ; the (information) entropy of  $p$  is given by the following function

---

<sup>2</sup>we tacitly assume that the normalization constraint  $\sum_{i=1}^n p_i = 1$  has been added as the observable  $\varphi_1^i = 1 \forall i \in \chi$ .

(entropy) defined on  $\mathbb{R}_+^n$

$$S(p) = - \sum_{i=1}^n p_i \ln p_i. \tag{12}$$

Note that  $S$  is a convex function and that the Hessian of  $S$  is

$$H_S(p) = -Diag\left[\frac{1}{p_i}\right].$$

We will call  $\mathbb{R}_+^n$  the space of microscopic states of the system and  $\mathbb{R}^k$  the space of macroscopic states.

Remark. In the sequel, the fact that the linear operator  $\varphi$  will be restricted to the manifold with boundary  $\mathbb{R}_+^n$  will play no role since all the critical points of  $S$  will be in the interior of  $\mathbb{R}_+^n$ , where  $T\mathbb{R}_+^n = T\mathbb{R}^n$ .

(M.E.P.) The selected probability distribution<sup>3</sup>  $p$  is the (unique) one *maximizing* the information entropy (uncertainty)

$$S(p) = - \sum_{i=1}^n p_i \ln p_i$$

among those verifying

$$\langle \varphi_L \rangle = \sum_{i=1}^n p_i \varphi_L^i = c_L, \quad L = 1, \dots, k \leq n. \tag{13}$$

Now we show that the M.E.P. procedure can be recast in the scheme developed in the last Section. (See also [9], for a different approach of M.E.P. in the framework of Symplectic Geometry). The above constrained maximization problem (M.E.P.) can be dealt with by the Lagrange Multipliers Method by introducing the function

$$\mathcal{S}(c; \lambda, p) := (\varphi(p) - c) \cdot \lambda - S(p).$$

Note that the apparatus of the Lagrange multipliers method supplemented by second order sufficient conditions of Proposition 1.3 and Proposition 1.4 determines the constrained *local* maximizers of the entropy function. The main feature of the linear constraint case is that there is a *unique* maximizer.

As a straightforward consequence of Proposition 1.3 (with  $m = p$  and  $G_{mm} = -H_S$ ) we have the following result:

**Proposition 2.1** *Since  $\varphi$  defines a (linear) fibration and  $-H_S(p) \in Sym^+(n)$  is a positive definite matrix, then*

*i) condition (1) holds, hence  $\mathcal{S}(c; \lambda, p)$  is a Morse family where*

---

<sup>3</sup>From now on we use the letter  $p$  to denote a probability distribution while the letter  $\mathbf{p}$  was used above to denote a covector at  $q$

ii) all the  $k + n$  parameters  $(\lambda, p)$  can be eliminated; hence

iii)  $\Lambda$  in (8) is a Lagrangian submanifold locally transversal to the fibers of  $\pi$ .

As we have seen in Proposition 1.2 the critical set  $\mathcal{E}$

$$\mathcal{E} = \{(c; \lambda, p) : \frac{\partial \mathcal{S}}{\partial p} = \varphi^T \lambda - dS(p) = 0, \frac{\partial \mathcal{S}}{\partial \lambda} = \varphi p - c = 0\} \quad (14)$$

is, locally at  $c$ , the graph of a map

$$g : W \subset \mathbb{R}^k \rightarrow \mathbb{R}^{k+n}, \quad c \mapsto g(c) = (\tilde{\lambda}(c), \tilde{p}(c))$$

such that

$$\text{graph}(g) = \mathcal{E} \cap \{(c; \tilde{\lambda}(c), \tilde{p}(c)) : c \in W\}.$$

Now, setting  $\tilde{\mathcal{S}}(c) = \mathcal{S}(c; g(c))$  – see (10)– we have that

$$\tilde{\mathcal{S}}(c) = \mathcal{S}(\tilde{p}(c)) = -S(\tilde{p}(c)) \quad (15)$$

and, by (3) and Proposition 1.2, that

$$\frac{\partial \tilde{\mathcal{S}}}{\partial c}(c) = \left[ \frac{\partial \mathcal{S}}{\partial c}(c, u) + \frac{\partial \mathcal{S}}{\partial u}(c, u) \frac{\partial g(c)}{\partial c} \right] \Big|_{u=g(c)} = \frac{\partial \mathcal{S}}{\partial c}(c, u) \Big|_{u=g(c)} = -\tilde{\lambda}(c)$$

since  $(c, g(c)) \in \mathcal{E}$ . Hence the projected submanifold  $\Lambda$  is, locally at  $W$ , the graph of  $d\tilde{\mathcal{S}}(c)$

$$\Lambda_S = \{(c, \frac{\partial \tilde{\mathcal{S}}}{\partial c}(c)) = (c, -\tilde{\lambda}(c)) : c \in W\}. \quad (16)$$

It is instructive to look at the explicit form of the local map  $g$  that allows for the elimination of the extra parameters  $u = (\lambda, p)$ . The map  $g$  has to be determined by the equations  $\frac{\partial \mathcal{S}}{\partial u} = 0$  in (14). More in detail (here summation over repeated indices is understood):

$$\begin{cases} \varphi_{Li} p_i = c_L, & l = 1, \dots, k, \\ \varphi_{Mi} \lambda_M - \frac{\partial \mathcal{S}}{\partial p_i}(p) = \varphi_{Mi} \lambda_M + 1 + \ln p_i = 0, & i = 1, \dots, n. \end{cases} \quad (17)$$

Note that we do not take into account explicitly the normalization constraint. From the second equation we get

$$p_i = p_i(\lambda) := e^{-1 - \varphi_{Mi} \lambda_M} \quad (18)$$

which is *injective* since  $\varphi^T$  is injective by hypothesis (11); hence we have succeeded to give the parameters  $p$  as a function of the  $\lambda$ . Moreover, by substituting (18) in the first line of (17) we get the following system of  $k$  equations in the  $k$  unknowns  $\lambda_L$ :

$$c_L = \varphi_{Li} e^{-1 - \varphi_{Mi} \lambda_M} = -\frac{\partial}{\partial \lambda_L} \mathcal{F}(\lambda) \quad (19)$$

where  $\mathcal{F} : \mathbb{R}^k \rightarrow \mathbb{R}$  is defined as follows:

$$\mathcal{F}(\lambda) = \sum_{i=1}^n e^{-1-\varphi_{Mi}\lambda_M}.$$

If the gradient map in (19) has an (at least local) inverse  $\lambda = \tilde{\lambda}(c)$ , then by substitution in (18) we get  $p = \tilde{p}(c)$  and we are done. The existence of a local inverse is granted –through the implicit function theorem– by Proposition 2.1 which is the reformulation of Proposition 1.3 to the M.E.P. setting. However, the existence can be checked directly given the explicit form of  $g$  as follows: a sufficient condition for *local* invertibility of (19) is that the matrix

$$(H_{\mathcal{F}}(\lambda))_{LN} := \frac{\partial c_L}{\lambda_N} = -\varphi_{Li}\varphi_{Ni}e^{-1-\varphi_{Mi}\lambda_M} = -\langle \varphi_L \varphi_N \rangle$$

be non-singular.

The global invertibility can be investigated using Proposition 1.4: a sufficient condition for *global* invertibility of the map (19) on a bounded convex domain is that the matrix  $H_{\mathcal{F}}(\lambda)$  be (positive or negative) definite. But, for  $u \in \mathbb{R}^k$ ,

$$\begin{aligned} -u^T \cdot H_{\mathcal{F}} u &= \sum_{L,N=1}^k \langle \varphi_L \varphi_N \rangle u_L u_N = \left\langle \sum_{L,N=1}^k \varphi_L \varphi_N u_L u_N \right\rangle = \\ &= \left\langle \left( \sum_{L=1}^k \varphi_L u_L \right)^2 \right\rangle = \sum_{i=1}^n [(\varphi^T u)_i]^2 p_i(\lambda) \geq 0 \end{aligned}$$

from which we get

$$u^T \cdot H_{\mathcal{F}} u = 0 \Leftrightarrow \varphi^T u = 0 \Leftrightarrow u = 0$$

hence  $-H_{\mathcal{F}} \in Sym^+$  and we have proved the global invertibility therefore  $\Lambda$  is *globally* transversal to the fibers of  $\pi$ .

Of course, an explicit formula for the inverse of the gradient map (19) does not exist.

Let us denote with  $\tilde{\lambda} = \tilde{\lambda}(c)$  the inverse of (19). Then the generating function in (15) of the projected lagrangian submanifold (16) has the form

$$\begin{aligned} \tilde{\mathcal{S}}(c) &= \mathcal{S}(\tilde{p}(c)) = S(p(\tilde{\lambda}(c))) = - \sum_{i=1}^n e^{-1-\varphi_{Mi}\tilde{\lambda}_M} \ln(e^{-1-\varphi_{Mi}\tilde{\lambda}_M}) \\ &= \mathcal{F}(\tilde{\lambda}) + \tilde{\lambda}_M \varphi_{Mi} p_i(\tilde{\lambda}) = \mathcal{F}(\tilde{\lambda}) + \tilde{\lambda}_M f_M. \end{aligned}$$

**Remark 1.** The above relation between  $\tilde{\mathcal{S}}(f)$  and  $\mathcal{F}(\tilde{\lambda})$  is the same as the one between the Lagrangian  $L(\dot{q})$  and the Hamiltonian  $H(p)$ .

**Remark 2.** Legendre transform. It is now easy to show that the inverse of a gradient map, if it exists, it is a gradient map too. In fact we have

$$\begin{aligned} \frac{\partial \tilde{\mathcal{S}}}{\partial c_L}(c) &= \frac{\partial \mathcal{F}}{\partial c_L}(\tilde{\lambda}(c)) + \frac{\partial}{\partial c_L}(\tilde{\lambda}_M c_M) = \frac{\partial \mathcal{F}}{\partial c_L}(\tilde{\lambda}(c)) + \tilde{\lambda}_L + \frac{\partial}{\partial c_L}(\tilde{\lambda}_M) c_M \\ &= \frac{\partial \mathcal{F}}{\partial c_L}(\tilde{\lambda}(c)) + \tilde{\lambda}_L + \frac{\partial}{\partial c_L}(\tilde{\lambda}_M) \left( -\frac{\partial \mathcal{F}(\tilde{\lambda})}{\partial \lambda_M} \right) = \tilde{\lambda}_L(c). \end{aligned} \tag{20}$$

**Remark 3.** It is interesting to consider the case that the matrix  $\varphi$  depends smoothly on  $\alpha$  real parameters  $(a_1, \dots, a_\alpha) \in A \subset \mathbb{R}^\alpha$ , that is

$$\varphi = \varphi(a), \quad rk\varphi(a) = k \quad \forall a \in A$$

As a consequence, the above Legendre transform formula inherits a dependence on the parameters  $a$

$$\tilde{\mathcal{S}}(c, a) = \mathcal{F}(\tilde{\lambda}(c, a), a) + \tilde{\lambda}_M(c, a) c_M.$$

### 3 Examples of M.E.P. with nonlinear constraints

In the case of M.E.P. with linear constraints we have seen that the Lagrangian submanifold associated to the constrained maximization problem is always (globally) transversal to the fibers of  $\pi$ , hence the M.E.P. singles out a *unique* statistical state. We will see that this feature is no longer conserved in case of M.E.P. with *nonlinear* constraints.

We point out some of the relevant features of the nonlinear case through some examples and remarks.

**Example 1.** Let  $\chi = \{1, \dots, n\}$  be the space state of a system and suppose that the observable  $\varphi$  satisfy the hypothesis : there exists  $i, j \in \chi$  such that  $\varphi_i > 0$ , and  $\varphi_j < 0$ . Then for every  $c > 0$  there are solutions  $p_i > 0$  to the constraint equation

$$\langle \varphi \rangle = \sum_{i=1}^n p_i \varphi_i = c \quad (+)$$

or

$$\langle \varphi \rangle = \sum_{i=1}^n p_i \varphi_i = -c. \quad (-)$$

Now we suppose that the result of an experiment gives the value of  $\langle \varphi \rangle^2$ ; therefore we are lead to find the maximum entropy probability distribution subject to the nonlinear constraint

$$\langle \varphi \rangle^2 - c^2 = 0.$$

The solutions to this constrained maximization problem are the union of the solution (which are unique) to the maximum entropy problem with linear constraints (+) and (-). These latter can be found as explained in Sect.2 and are

$$p_i^{(+)} = e^{-1-\lambda^{(+)}\varphi_i}, \quad p_i^{(-)} = e^{-1-\lambda^{(-)}\varphi_i}, \quad i = 1, \dots, n.$$

Hence, for every  $c > 0$  there are exactly two maximum entropy distributions, and the critical set is

$$\mathcal{E} = \bigcup_{c>0} \{(c, p^{(+)}(c), \lambda^{(+)}(c), p^{(-)}(c), \lambda^{(-)}(c))\}.$$

The associated Lagrangian submanifold is given by equation

$$\Lambda = \{(c, \mathbf{p}) \in T^*\mathbb{R} : \mathbf{p} = \lambda^{(+)}(c), \text{ or } \mathbf{p} = \lambda^{(-)}(c)\}$$

It is easy to see that  $\Lambda$  has the structure of the graph of a two-valued function defined for  $c > 0$ . Note that for  $c = 0$  the two constraints (+) and (−) do coincide, therefore the above function is single-valued. However, for  $c = 0$  the constraint  $\zeta(p) = \langle \varphi \rangle^2$  fails to have maximal rank on the set of solutions of  $\zeta(p) = c = 0$ . The following ones are physical examples that can be easily recast into this scheme.

**Example 2.** A physical instance of the M.E.P. with nonlinear constraints is the Curie–Weiss approximation of the Ising model for ferromagnetism that we briefly describe below (see e.g.[4]). Let  $\mathcal{L} \subset \mathbb{Z}^2$  be a lattice containing  $N$  sites each occupied by an atom whose spin can assume the value  $s = +1$  or  $s = -1$ . The set of possible spin configurations of the lattice is  $\Omega = \{-1, +1\}^{\mathcal{L}}$ . The potential energy of the lattice (ferromagnet) in the configuration  $\omega \in \Omega$  and subject to an external magnetic field  $B$  is defined as

$$H(\omega, B) = -\frac{1}{2} \sum_{[i,j]} s_i(\omega) s_j(\omega) - B \sum_{i=1}^N s_i(\omega) \quad (21)$$

where  $[i, j]$  is a couple of nearest-neighborhood sites. We are therefore in the case of an observable (the energy) depending on a parameter (the field strength  $B$ ). The Curie–Weiss approximation of this model is to neglect the interactions between neighboring spins and hence to assume that the values of the spin at the different sites  $i = 1, \dots, n$  are independent identically distributed random variables taking values in  $\{-1, +1\}$ ; therefore

$$\langle s_i \rangle = \langle s_j \rangle \quad \forall i, j = 1, \dots, N, \quad \text{and} \quad \langle s_i s_j \rangle = \langle s_i \rangle \langle s_j \rangle.$$

The average value of the energy of the lattice is

$$\begin{aligned} \langle H \rangle &= \sum_{\omega \in \Omega} p(\omega) H(\omega) = \sum_{\omega \in \Omega} p(\omega) \left( -\frac{1}{2} \sum_{[i,j]} s_i(\omega) s_j(\omega) - B \sum_{i=1}^N s_i(\omega) \right) \\ &= -\frac{1}{2} \sum_{[i,j]} \sum_{\omega \in \Omega} p(\omega) s_i(\omega) s_j(\omega) - B \sum_{i=1}^N \sum_{\omega \in \Omega} p(\omega) s_i(\omega) = \\ &= -\frac{1}{2} \sum_{[i,j]} \langle s_i s_j \rangle - BN \langle s \rangle = -N \langle s \rangle^2 - BN \langle s \rangle. \end{aligned}$$

where

$$\langle s \rangle = \langle s_i \rangle = p_+(1) - p_-(-1) = p_+ - p_-.$$

Here  $p_+, p_-$  is the probability of the spin up, spin down configuration respectively for the spin at an arbitrary site. Therefore the linear constraint on the lattice energy

$$\langle H \rangle = E$$

where  $H_{min} \leq E \leq H_{max}$ , is equivalent to the *nonlinear* constraint on  $\langle s \rangle = p_+ - p_-$

$$-\langle s \rangle^2 - B \langle s \rangle = \frac{E}{N} =: \gamma \quad (22)$$

for the average value of the spin on a single site.

From now on we consider the space state to be  $\{1, -1\}$  and we look for the probability distribution  $p = (p_+, p_-) \in \mathbb{R}_+^2$  maximizing the entropy (12) subject to the normalization and energy constraints. In order to apply the theory developed in the previous Sections we are lead to introduce the normalization constraint in the general form  $\sum p_i = p_+ + p_- = \alpha > 0$ ; in this setting the average value of the spin is

$$\langle s \rangle = \frac{1}{\alpha}(p_+ - p_-), \quad \langle s \rangle \in [-1, 1].$$

The constraint equation are therefore, see (22)

$$\zeta_1(p) = p_+ + p_- = \alpha > 0, \tag{23}$$

$$\zeta_2(p) = -\frac{1}{\alpha^2}(p_+ - p_-)^2 - \frac{1}{\alpha}B(p_+ - p_-) = \gamma. \tag{24}$$

with

$$\text{rk } d\zeta(p) = \max = 2 \quad \text{for } p_+ - p_- \neq -\frac{\alpha B}{2}.$$

In this rather special example (with  $k = n$ , the number of constraints equal to the dimension of the microscopic states manifold) the fiber  $\zeta^{-1}(c)$ ,  $c = (\alpha, \gamma)$  is made of *isolated points*; these are the intersections in the positive quarter plane  $(p_+, p_-) \in \mathbb{R}_+^2$  of the line  $\zeta_1(p) - \alpha = 0$  with the parabola  $\zeta_2(p) - \gamma = 0$ . There can be none, one or two intersection points.

**Example 3.** Planar Potts model. A straightforward generalization of the Ising model is the planar Potts model. Here the spin vector can point in  $q \geq 2$  directions in the plane, each forming an angle  $\theta_l = \frac{2\pi l}{q}$ ,  $l = 1, \dots, q$  with a fixed reference axis in the plane. The set of possible spin configurations of the lattice is  $\Omega = \{1, \dots, q\}^{\mathcal{L}}$ . The potential energy of the lattice (ferromagnet) in the configuration  $\omega \in \Omega$  and subject to an external magnetic field  $B$  is

$$\begin{aligned} H(\omega, B) &= -\frac{1}{2} \sum_{[i,j]} s_i(\omega) \cdot s_j(\omega) - B \sum_{i=1}^N s_i(\omega) \cdot n \\ &= -\frac{1}{2} \sum_{[i,j]} \cos(\theta_i(\omega) - \theta_j(\omega)) - B \sum_{i=1}^N \cos(\theta_i(\omega)) \end{aligned}$$

where  $[i, j]$  is a couple of nearest-neighborhood sites, dot denotes scalar product and  $n$  is the versor of the fixed reference axis. In the Curie-Weiss approximation of statistical independence of the state  $s_i$  with respect to  $s_j$  (i.e. there is no correlation between the spin at neighboring sites) the average value of the lattice energy is given by

$$\langle H \rangle = -Np^T A p - N B u \cdot p = E \tag{25}$$

where the superscript  $T$  denotes transposition,  $A \in \text{Sym}(q)$  is the interaction symmetric matrix

$$A_{lm} = \cos(\theta_l - \theta_m) = \cos\left(\frac{2\pi}{q}(l - m)\right), \quad l, m = 1, \dots, q$$

and  $u \in \mathbb{R}^q$  has components

$$u_l = \cos(\theta_l) = \cos\left(\frac{2\pi l}{q}\right), \quad l = 1, \dots, q.$$

The spectrum of the matrix  $A$  has the following expression

$$Spec(A) = \{0, \dots, 0, \frac{q}{2}, \frac{q}{2}\}$$

therefore equation (25) defines an hypercylinder in  $\mathbb{R}^q$ . The fiber  $\zeta^{-1}(c)$ , ( $c = \alpha, \frac{E}{N}$ ) is the intersection, in the positive octant, of the normalization constraint hyperplane

$$\sum_{i=1}^q p_i = \alpha$$

with the (convex) energy hypercylinder (25).

#### 4 Phase transitions in the M.E.P. setting

We consider, as in the previous examples, a physical system in which, for simplicity' sake, the only observable of interest is the energy, and we suppose that the information we have on the system energy  $E$  places a (possibly nonlinear) constraint on the probability distribution

$$\zeta(p) - E = 0, \quad \text{where } \zeta : \mathbb{R}^n \rightarrow \mathbb{R}.$$

Then the set of *macroscopic* states of the system, determined by the M.E.P., is the Lagrangian submanifold –see (8)–

$$\Lambda = \{(E, -\beta) \in T^*\mathbb{R} : (E, \beta, p) \in \mathcal{E}\}$$

where  $\beta$  is the Lagrangian multiplier associated to the energy constraint that can be given in the linear case –see (20) and (15)–as

$$\beta = \frac{\partial \tilde{\mathcal{S}}(E)}{\partial E} = -\frac{\partial S(\tilde{p}(E))}{\partial E}. \tag{26}$$

In case  $\Lambda$  is *transversal* to the base manifold, then there is a *unique*  $\beta$  associated to  $E$ ; if  $\Lambda$  is in general position w.r.t. the base manifold, then there can be multiple values  $\beta$  associated to the system energy  $E$ . This multi-valuedness is interpreted as a phase transition for the system. Note that, as developed in the previous Sections, phase transitions are possible within this scheme only if *nonlinear* constraints are in force.

Critical values of  $E$ , where transitions can occur, can be (in principle) determined through formula (4). However, (4) requires the computation of the solution of a nonlinear system of  $(n + k) \approx n$  equations and of the determinant of a order  $\approx n$  matrix. Therefore (4) gives both an experimentally testable quantity (the critical values of  $E$  or the related temperature  $\beta$ ) and sets the limits of applicability of this formalism in terms of system complexity.

Finally, let us mention briefly some other approaches to the existence of phase transitions for discrete systems with finite volume (i.e. not in the thermodynamic limit).

For some Hamiltonian systems the entropy of the system having energy  $E$  can be, explicitly or at least numerically, computed through the Boltzmann formula

$$S(E) = k \ln W(E)$$

where  $W(E)$  is the area in phase space of the  $E$ -level set of the Hamiltonian  $H(q, p)$ ; the system temperature is defined as (compare with the above formula (26))

$$\beta = \frac{1}{kT} = \frac{\partial S(E)}{\partial E}.$$

If  $S(E)$  is not a convex function of  $E$  on its domain, i.e.

$$-\frac{\partial^2 S(E)}{\partial^2 E} = \frac{1}{T^2 C_v} \notin Sym^+,$$

where  $C_v$  is the specific heat at constant volume, then there may be multiple values of the energy for a given value of the temperature, which is again interpreted with the presence of a phase transition for the system.

This approach (which dates back to Boltzmann) has been applied in literature to self-gravitating system ([10], [11]), to the Potts models ([7]), and to mechanical systems with non convex potential energy ([1], [5]). For systems with an Hamiltonian of the form kinetic plus potential energy, the phenomenon of *negative* specific heat stems from the fact that in some energy intervals, an increase of the total energy  $E$  determines an increase of the average system' potential energy and a diminution of the average kinetic energy (i.e. the temperature of the system).

## Acknowledgments

The author wishes to thank F. Cardin for useful discussion on the subject of this paper, P. Dai Pra for suggesting the Potts model example and C. Zanella for advices in the Proof of Proposition 1.3.

## References

- [1] V.L. Berdichevsky *Thermodynamics of chaos and order* Pitman Monographs and Surveys in Pure and Applied Mathematics 90 Addison Wesley Longman
- [2] M. Berger and M. Berger *Perspectives in Nonlinearity. An Introduction to Nonlinear Analysis* W. A. Benjamin Inc. New York, (1968)
- [3] D. P. Bertsekas *Constrained Optimization and Lagrange Multipliers Methods* Athena Scientific, Belmont, Massachusetts, (1996)

- [4] R. Brout *Phase Transitions*, in: Statistical Physics. Phase transitions and Superfluidity. vol. 1 pp. 5-103 Brandeis University Summer Institute in Theoretical Physics, Gordon and Breach Science Publishers (1966)
- [5] F. Cardin, M. Favretti *On the Helmholtz–Boltzmann Thermodynamics of Mechanical Systems* Continuum Mech. Thermodyn. 16, 15–29 (2004)
- [6] M. Favretti *Isotropic submanifold generated by the Maximum Entropy principle and Onsager reciprocity relations* Journal of Functional Analysis, to appear.
- [7] D.H.E. Gross, A. Ecker and X.Z. Zhang *Microcanonical thermodynamics of first order phase transitions studied in the Potts model* Ann. Physik (8) 5, no. 5, 446–452 (1996), available at <http://xxx.lanl.gov/abs/cond-mat/9607150>
- [8] E. T. Jaynes *Information theory and statistical mechanics* Phys. Rev. 106, 620–630 (1957)
- [9] S. Omohundro *Geometric Perturbation Theory in Physics*, World Scientific Publishing Co., Singapore, (1986)
- [10] W. Thirring *Systems with Negative Specific Heat* Z. Phys. 235, 339–352 (1970)
- [11] E.V. Votyakov, H.I. Hidmi, A. De Martino and D.H.E. Gross *Microcanonical mean–field thermodynamics of self–gravitating and rotating systems* available at <http://arXiv.org/abs/cond-mat/0202140> (2002)
- [12] A. Weinstein, *Lectures on symplectic manifolds* C.B.M.S. Conferences Series in Mathematics, A.M.S. 29 (1977)

©2005 by MDPI (<http://www.mdpi.org>). Reproduction for noncommercial purposes permitted.

