PROCEEDINGS OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 43, Number 1, March 1974

A UNIFIED TREATMENT OF SOME THEOREMS ON POSITIVE MATRICES

GÉRARD LETAC

ABSTRACT. Various theorems on positive matrices are shown to be corollaries of one general theorem, the proof of which bears on Legendre functions, as used in Rockafellar's *Convex analysis*.

1. **Introduction: The main theorem.** Let X be a finite set, $(\mu(x))_{x\in X}$ strictly positive numbers, and H a fixed linear subspace of $\boldsymbol{R}^X$. We shall prove the following:

THEOREM 1. *There exists a unique (nonlinear) map from $\boldsymbol{R}^X$ to H, denoted $f\mapsto h_f$, such that*

$$\sum_{x\in X}[\exp f(x) - \exp h_f(x)]g(x)\mu(x) = 0$$

for all g in H.

2. **A first application: Matrices with prescribed marginals.** Given an n-sequence $r=(r_i)_{i=1}^n$ and an m-sequence $s=(s_j)_{j=1}^m$ of nonnegative numbers such that $\sum_{i=1}^n r_i=\sum_{j=1}^m s_j$, denote by $\mathcal{M}(r,s)$ the set of (n,m) matrices (a_{ij}) with $a_{ij}\geqq 0$ such that $r_i=\sum_{j=1}^m a_{ij}$ and $s_j=\sum_{i=1}^n a_{ij}$ for all $i=1,2,\cdots,n$ and $j=1,2,\cdots,m$. Also, let $E=\{1,2,\cdots,n\}$ and $F=\{1,2,\cdots,m\}$. We define the linear map c from $\boldsymbol{R}^E\oplus\boldsymbol{R}^F$ to $\boldsymbol{R}^{E\times F}$ by:

$$c[(b_i)_{i\in E},(b'_j)_{j\in F}] = (b_i+b'_j)_{(i,j)\in E\times F}.$$

If X is a subset of $E\times F$, π denotes the canonical map from $\boldsymbol{R}^{E\times F}$ to $\boldsymbol{R}^X$, i.e.

$$\pi[(a_{ij})_{(i,j)\in E\times F}] = (a_{ij})_{(i,j)\in X}.$$

We say also that X is an (r,s) *pattern* if there exists (a_{ij}) in $\mathcal{M}(r,s)$ such that $X=\{(i,j);\,a_{ij}>0\}$.

Now the first corollary to the Theorem 1 is:

COROLLARY 1. *Let two sequences $r=(r_i)_{i=1}^n$ and $s=(s_j)_{j=1}^m$ of nonnegative numbers with $\sum_{i=1}^n r_i=\sum_{j=1}^m s_j$, M an (n,m) matrix with $\mu_{ij}\geqq 0$ and $X=\{(i,j);\,\mu_{ij}>0\}$.*

Received by the editors February 6, 1973.

AMS (MOS) subject classifications (1970). Primary 15A48; Secondary 58C20, 94A15.

Key words and phrases. Positive matrices, entropy, gradient.

Then there exist two diagonal matrices

$$D_b = (e^{b_1}, e^{b_2}, \cdots, e^{b_n})$$

and

$$D'_{b'} = (e^{b'_1}, \cdots, e^{b'_m})$$

with positive diagonal elements such that $D_bMD'_{b'}$ is in $\mathscr{M}(r, s)$ if and only if X is an (r, s) pattern. Furthermore, the set (b, b') of $\boldsymbol{R}^E \oplus \boldsymbol{R}^F$, such that $D_bMD'_{b'}$ is in $\mathscr{M}(r, s)$, is exactly a translate of $\mathrm{Ker}[\pi \circ c]$ *if nonempty.*

PROOF. If b and b' exist, the fact that X is an (r, s) pattern is obvious. Conversely, suppose that X is an (r, s) pattern. Then there exists $(a_{ij}) \in \mathscr{M}(r, s)$ with $X = \{(i, j); a_{ij} > 0\}$. Denote $f_{ij} = \mathrm{Log}(a_{ij}/\mu_{ij})$ when $(i, j) \in X$ and apply the theorem to this $f \in \boldsymbol{R}^X$ and to H, the range in $\boldsymbol{R}^X$ of $\pi \circ c$. Then there exists $h = (h_{ij})_{(i,j)\in X}$ in H such that

$$\sum_{(i,j)\in X} a_{ij} g_{ij} = \sum_{(i,j)\in X} \exp(h_{ij}) g_{ij} \mu_{ij}$$

for all $(g_{ij})_{(i,j)\in X}$ of H, and such h is unique. Writing now $h_{ij} = b_i + b'_j$ for some $(b_i)_{i=1}^n \in \boldsymbol{R}^E$ and $(b'_j)_{j=1}^m \in \boldsymbol{R}^F$, we have $(e^{b_i}\mu_{ij}e^{b'_j})_{(i,j)\in E} \in \mathscr{M}(r, s)$.

All suitable (b, b') must satisfy $\pi \circ c(b, b') = h$, and this ends the proof.

In order to complete Corollary 1 we have to specify $\mathrm{Ker}[\pi \circ c]$ for a given $X \subset E \times F$. We index E and F such that $E \cap F = \varnothing$. We consider the linear graph (nonoriented) with $E \cup F$ as the set of vertices and X as the set of edges. The connected components of that linear graph can be written $(E_\alpha \cup F_\alpha)_{\alpha\in A}$, where A is some index set (with $E_\alpha \cup F_\alpha$ nonempty, but E_α or $F_\alpha = \varnothing$ can happen), and we call $(E_\alpha)_{\alpha\in A}$ and $(F_\alpha)_{\alpha\in A}$ the partitions of E and F associated to X.

PROPOSITION 1. *Let X be a subset of $E \times F$, π the canonical map from $\boldsymbol{R}^{E\times F}$ to $\boldsymbol{R}^X$, $(E_\alpha)_{\alpha\in A}$ and $(F_\alpha)_{\alpha\in A}$ the associated partitions of E and F. Then* $\dim[\mathrm{Ker}(\pi \circ c)]$ *= number of α such that $E_\alpha \times F_\alpha$ is not empty. Furthermore, when (i_1, j_1) and (i_2, j_2) are in $E_\alpha \times F_\alpha$ and $(b, b') \in$* $\mathrm{Ker}[\pi \circ c]$, *then*

$$b_{i_1} = b_{i_2} = -b_{j_1} = -b_{j_2}.$$

PROOF. Suppose $E_\alpha \times F_\alpha$ is nonempty; take (i_1, j_1) and (i_2, j_2) in $E_\alpha \times F_\alpha$ and (b, b') in $\mathrm{Ker}(\pi \circ c)$. From the definition of $E_\alpha \cup F_\alpha$, either there exists a sequence of X

$$(e_0, f_0), (e_1, f_0), (e_1, f_1), (e_2, f_1), \cdots, (e_n, f_{n-1})$$

or there exists a sequence of X

$$(e_0, f_0), (e_0, f_1), (e_1, f_1), (e_1, f_2), \cdots, (e_n, f_n)$$

with $e_0 = c_1$ and $e_n = c_2$.

Since $b_e+b'_f=0$ when $(e,f)\in X$ this implies that $b_{i_1}=b_{i_2}$. In the same way we prove $b'_{j_1}=b'_{j_2}$. To see that $b_{i_1}=-b'_{j_1}$, we find a path from i_1 to j_1 in an analogous manner. Then dim $\operatorname{Ker}[\pi\circ c]$ is not larger than the number of nonempty $E_\alpha\times F_\alpha$. From this, it is easy to finish the proof.

REMARKS. Corollary 1 was conjectured in 1964 by P. Thionet who is a French statistician [**10**], and in a weaker form in 1960 [**9**]. Proofs in particular cases are given from [**1**]. Independently, the case $n=m$, $r_i=s_j=1$ was intensively studied in a paper by R. Sinkhorn in 1964 [**7**]; see [**3**] for a recent proof and references.[1]

It is easy to prove the following: X is an (r,s) pattern if and only if for any $x\in X$ there exists an (r,s) pattern $Y(x)$ such that $x\in Y(x)\subset X$. An interesting feature of the case of doubly stochastic matrices is that the patterns $Y(x)$ are associated in a natural way to permutation matrices, permitting a characterisation, via a König theorem, of a corresponding pattern as a direct sum of so-called fully indecomposable matrices [**3**]. It would be interesting to get such a combinatorial characterisation for a general (r,s) pattern.

3. **Second application: A theorem due to D. J. Hartfiel.** Let $E=\{1,\cdots,n\}$ and $\mathcal{M}_s$ be the set of n-square matrices (a_{ij}) with $a_{ij}\geqq 0$ and $\sum_{j=1}^n a_{ij}=\sum_{j=1}^n a_{ji}$ for all i in E. We define the linear map d from $\boldsymbol{R}^E$ to $\boldsymbol{R}^{E\times E}$ by:

$$d[(b_i)_{i\in E}]=(b_i-b_j)_{(i,j)\in E\times E}.$$

If X is a subset of $E\times E$, π is the canonical map from $\boldsymbol{R}^{E\times E}$ to $\boldsymbol{R}^X$. We say also that X is a *symmetric* pattern if there exists (a_{ij}) in $\mathcal{M}_s$ such that $X=\{(i,j);a_{ij}>0\}$. The second corollary of Theorem 1 is essentially in D. J. Hartfiel [**2**].

COROLLARY 2. *Let $M=(\mu_{ij})$ be an n-square matrix with $\mu_{ij}\geqq 0$ and $X=\{(i,j);\mu_{ij}>0\}$. Then there exists one diagonal matrix $D_b=(e^{b_1},e^{b_2},\cdots,e^{b_n})$ with positive diagonal elements such that $D_bMD_b^{-1}$ is in $\mathcal{M}_s$, if and only if X is a symmetric pattern. Furthermore, the set of $b\in\boldsymbol{R}^E$ such that $D_bMD_b^{-1}$ is in $\mathcal{M}_s$ is exactly a translate of* $\operatorname{Ker}[d\circ\pi]$ *if nonempty.*

PROOF. Easy, taking $H=\pi\circ d[\boldsymbol{R}^E]$.

If we want now to complete the Corollary 2, we have to characterise $\operatorname{Ker}[\pi\circ c]$ for a given $X\subset E\times E$ and characterise symmetric patterns.

[1] NOTE ADDED IN PROOF. Professor Caussinus from the University of Toulouse has indicated to the author that a complete story of Corollary 1 can be found in Chapter 4 of M. Bacharach's monograph, *Biproportional matrices and input-output change*, Cambridge University Press, 1970. The first proof of Corollary 1 must be credited to W. M. Gorman (1963).

We consider the oriented graph with E as the set of vertices and X as the set of arcs.

PROPOSITION 2. *X is a symmetric pattern if and only if for any $e \in E$ there exists a sequence $e_0, e_1, \cdots, e_n$ in E such that $e_0=e_n=e$ and $(e_{i-1}, e_i) \in X$ for $i=1, \cdots, n-1$ $(n \geqq 1)$ (in terms of the graph (E, X) every vertex belongs to a circuit).*

Now we suppose that X is a symmetric pattern and we partition the set of vertices of the graph (E, X) into connected components $(E_\alpha)_{\alpha \in A}$ where A is some index set and $E_\alpha \neq \varnothing$. These components correspond to "irreducible matrices" of **[2]**. We say that E_α is the partition of E associated to the symmetric pattern X.

PROPOSITION 3. *Let X be a symmetric pattern of $E \times E$, $(E_\alpha)_{\alpha \in A}$ the associated partition of E. Then* $\dim[\operatorname{Ker}(\pi \circ d)]$ *= number of elements of A. Furthermore $b \in \operatorname{Ker}(\pi \circ d)$ and i_1 and $i_2 \in E_\alpha$ implies $b_{i_1} = b_{i_2}$.*

PROOF. Similar to Proposition 1.

4. **Other applications.** Following the same vein, we may now consider the result of R. Sinkhorn **[8]**, stating that if M is an (n, n) symmetric matrix with strictly positive coefficients there exists one diagonal matrix D with positive diagonal coefficients such that $DMD \in \mathscr{M}(r, r)$, where $r=(r_i)_{i=1}^n$ is given (Marcus and Newman **[4]** consider the doubly stochastic case).

Denote $E=\{1, \cdots, n\}$ and define the linear map e from $\boldsymbol{R}^E$ to $\boldsymbol{R}^{E \times E}$ by:

$$e[(b_i)_{i \in E}] = (b_i + b_j)_{(i,j) \in E \times E}.$$

If X is a subset of $E \times E$, call X an *r-pattern* if there exists a symmetric matrix (a_{ij}) of $\mathscr{M}(r, r)$ such that $X=\{(i, j); a_{ij} > 0\}$.

COROLLARY 3. *Let $M=(\mu_{ij})$ be a symmetric matrix with $\mu_{ij} \geqq 0$, $r=(r_i)_{i=1}^n$ a sequence such that $r_i \geqq 0$ and $X=\{(i, j); \mu_{ij} > 0\}$. Then there exists one diagonal matrix $D_b=(e^{b_1}, \cdots, e^{b_n})$ with positive diagonal elements such that $D_b M D_b$ is in $\mathscr{M}(r, r)$ if and only if X is an r-pattern. Furthermore, the set of $b \in \boldsymbol{R}^E$ such that $D_b M D_b$ is in $\mathscr{M}(r, r)$ is exactly a translate of* $\operatorname{Ker}[e \circ \pi]$ *if nonempty.*

PROOF. Easy, taking $H = r \circ e[\boldsymbol{R}^E]$.

A similar study of dimension of $\operatorname{Ker}[\pi \circ e]$ can be done considering the number of components of the graph (E, X); characterisation of r-patterns depends on r in an unknown way, except in the doubly stochastic case, which is easy to work.

A trivial application of Theorem 1 is conditional expectation on a finite probability space: Let $(X_\alpha)_{\alpha\in A}$ be a partition of X; consider the boolean algebra $\mathfrak{A}$ generated by $(X_\alpha)_{\alpha\in A}$ and take H as space of $\mathfrak{A}$-measurable functions.

5. **Proof of Theorem 1.** Consider the function F defined on H by:

$$F(h) = \sum_{x\in X} \mu(x)\exp h(x).$$

Since $\mu(x)>0$ for all x, F is strictly convex. One can easily check that F is continuously differentiable and that grad F_h, element of the dual space H^* of H, is given by:

$$\text{grad } F_h(g) = \sum_{x\in X} \mu(x)g(x)\exp h(x)$$

for all g in H. Now to each f of $\boldsymbol{R}^X$, we associate one point h_f^* of H^* defined by:

$$h_f^*(g) = \sum_{x\in X} \mu(x)g(x)\exp f(x) \quad \text{for all } g \in H.$$

We have to show that there exists one and only one point $h_f \in H$ such that $h_f^* = \text{grad } F_{h_f}$.

Now we introduce the conjugate function F^* of F, valued in $\boldsymbol{R}\cup\{+\infty\}$ and defined on H^* by:

$$F^*(h^*) = \sup_{h\in H} [h^*(h) - F(h)].$$

We denote by C^* the interior of $\{h^*; F^*(h^*)<\infty\}$. We use the following inequality:

$$be^a - e^b \leqq a(e^a - 1) \qquad (a, b \in \boldsymbol{R})$$

(to see this, replace b with $a+b$) to get $F^*(h_f^*) \leqq \sum_{x\in X} \mu(x)f(x)(e^{f(x)}-1)$ for all f in $\boldsymbol{R}^X$. Clearly the range of $f\mapsto h_f^*$ in H^* is an open set, so $h_f^* \in C^*$. We made an appeal to Rockafellar [**6**, p. 258, Theorem 26.5]: we have shown that (H, F) and (C^*, F^*) are convex functions of Legendre type; from the quoted theorem we can assert that the map $h\mapsto \text{grad } F_h$ from H to H^* is injective and that the range of this map is C^*; this ends the proof.

6. **Other results and remarks.** The trivial example of conditional expectations quoted at the end of §4 leads us to ask if $P_H:\boldsymbol{R}^X\to H$ defined by $P_H f = h_f$ by means of Theorem 1 has some properties of ordinary linear projection.

THEOREM 2. *If $P_H:\boldsymbol{R}^X\to H$ is defined by $P_H f=h_f$ and if H and H' are two linear subspaces of $\boldsymbol{R}^X$, such that $H\subset H'$, then $P_H = P_H P_{H'}$.*

PROOF. Define $G_H:\boldsymbol{R}^X \to H^*$ by $(G_H f)(g)=\sum_{x\in X} e^{f(x)}g(x)\mu(x)$ and grad $F_H: H\to H^*$ as in the proof of Theorem 1. We know that grad $F_H \circ G_H = G_H$ and we have to prove

$$\text{grad } F_H \circ P_H \circ P_{H'} = G_H, \quad \text{or} \quad G_H \circ P_{H'} = G_H,$$

which is trivially true when $H \subset H'$.

From now, H is fixed and we denote $\sum_{x\in X} f(x)\mu(x)$ by $\int f\, d\mu$. Notations are those of Theorem 1.

THEOREM 3. *If H contains constant functions of X, then*:

$$\int e^f f\, d\mu \geqq \int e^{h_f} h_f\, d\mu$$

and the equality holds only if f is in H.

PROOF. From the inequality $t \geqq 1-e^{-t}$, for t in $\boldsymbol{R}$ (strict if $t\neq 0$), we get

$$\int e^f(1 - e^{h_f - f})\, d\mu \leqq \int e^f(f - h_f)\, d\mu.$$

Since H contains constants, the first member of the inequality is 0 (let $g=1$ in the statement of Theorem 1). Now $\int e^f h_f\, d\mu = \int e^{h_f} h_f\, d\mu$ (let $g=h_f$), and we are done. Case of equality is clear.

THEOREM 4. *Let h in H and $T: X\to X$. Suppose that either $h\circ T$ is in H or H contains constants and $\mu\circ T=\mu$. Then $h_{h\circ T}=h$ implies $h=h\circ T$.*

PROOF. If $h\circ T$ is in H, we use uniqueness of h_f in Theorem 1. In the other case, we apply Theorem 2 to $f=h\circ T$. But, since $\mu\circ T=\mu$, we have

$$\int e^f f\, d\mu = \int e^h h\, d\mu;$$

this is the case of equality of Theorem 3.

REMARKS. Choose $X=\{1,\cdots,n\}\times\{1,\cdots,m\}$, $\mu(x)\equiv 1$ and H as in §2. If f in $\boldsymbol{R}^X$ is such that $\int e^f\, d\mu = 1$, Theorem 3 gives an inequality about entropy of the joint law of probability, well known in information theory (see [5, pp. 146–157]). If T is a permutation of X, of course $\mu\circ T = \mu$; a nice interpretation of the second part of Theorem 4 is as follows.

Let $r=(\exp b_i)_{i=1}^n$ and $s=(\exp b'_j)_{j=1}^m$ such that

$$\sum_{i=1}^n \exp b_i = \sum_{j=1}^n \exp b'_j = 1.$$

Put the nm numbers $\exp(b_i+b'_j)$ on slips of paper, try to put them in an (n, m) matrix in order to get an element of $\mathscr{M}(r, s)$; you always get the matrix $(\exp(b_i+b'_j))$, that is to say, the product distribution of probability. Of course we need only Theorem 3, not Theorem 1, to get this result, which can be generalised to a product of countable spaces, with distributions r and s such that $\sum_i e^{b_i}b_i<\infty$, $\sum_j e^{b'_j}b'_j<\infty$. I think this is not true if $\sum_i e^{b_i}b_i=\infty$, but I know of no counterexamples.

Considering the exponential in Theorem 1 suffices for applications, but we could replace the exponential by any convex function φ positive and increasing, with primitive ψ. The proof of Theorem 1 would start from $F(h)=\sum_x \psi(h(x))\mu(x)$.

Generalisations of Theorem 1 to measured spaces (X, μ) will be done in a forthcoming paper.

BIBLIOGRAPHY

1. H. Caussinus, *Contribution à l'analyse statistique des tableaux de corrélation*, Ann. Fac. Sci. Univ. Toulouse (4) **29** (1965), 77–182. MR **39** #3672.

2. D. J. Hartfiel, *Concerning diagonal similarity of irreducible matrices*, Proc. Amer. Math. Soc. **30** (1971), 419–425. MR **43** #7446.

3. D. London, *On matrices with a doubly stochastic pattern*, J. Math. Anal. Appl. **34** (1971), 648–652. MR **43** #7448.

4. M. Marcus and M. Newman, *The permanent of a symmetric matrix*, Notices Amer. Math. Soc. **8** (1961), p. 595.

5. A. Renyi, *Foundations of probability*, Holden-Day, San Francisco, Calif., 1970. MR **41** #9314.

6. R. T. Rockafellar, *Convex analysis*, Princeton Math. Series, no. 28, Princeton Univ. Press, Princeton, N.J., 1970. MR **43** #445.

7. R. Sinkhorn, *A relationship between arbitrary positive matrices and doubly stochastic matrices*, Ann. Math. Statist. **35** (1964), 876–879. MR **28** #5072.

8. ———, *A relationship between arbitrary positive matrices and stochastic matrices*, Canad. J. Math. **18** (1966), 303–306. MR **32** #7574.

9. P. Thionet, *Sur certaines variantes des projections du tableau d'échanges inter-industriels*, Bull. Inst. Internat. Statist. **40** (1963), 431–446.

10. ———, *Note sur le remplissage d'un tableau à double entrée*, J. Soc. Statist. Paris, n° 10, 11, 12, (1964), 228–247.

DÉPARTEMENT DE MATHÉMATIQUES APPLIQUÉES, UNIVERSITÉ DE CLERMONT, B.P. N° 45 63170 AUBIÈRE, FRANCE

Current address: U.E.R. de Mathématiques, Université Paul-Sabatier, 118 Route de Narbonne, 31400 Toulouse, France