

NOTE ON THE UNIVERSAL BASES OF A PAIR OF POLYMATROIDS

Kazuo Murota
University of Tokyo

(Received February 19, 1988; Revised April 21, 1988)

Abstract This note gives a characterization of the universal pair of bases of a pair of polymatroids as the nearest pair of bases with respect to a class of pseudo-distances including the Kullback-Leibler divergence.

N. Megiddo considered the lexico-optimal flow problem in a multiterminal network $N = (V, A, c; S^+, S^-)$ (V : vertex set, A : arc set, c : capacity, S^+ : supply (source) vertices, S^- : demand (sink) vertices), which is to find a maximal flow such that the supply flow $(s^+(\nu)|\nu \in S^+)$ [resp., the demand flow $(s^-(\nu)|\nu \in S^-)$] is as proportional as possible to a given weight vector. This problem is treated by S. Fujishige as a special case of the lexico-optimal base problem for a single polymatroid.

This paper considers the problem of finding a maximal flow such that the supply flow s^+ and the demand flow s^- are as "near" as possible (where a one-to-one correspondence between S^+ and S^- is assumed to be given), and generalizes it to the problem of finding a "nearest" pair of bases of a pair of polymatroids. It is shown that the "nearest" pair coincides with the universal pair if either of the following criteria is adopted.

- (1) The f -divergence (a generalization of the Kullback-Leibler divergence) between the bases should be minimized;
- (2) The vector consisting of the ratios of the corresponding components of the bases should be lexicographically maximized.

1. Introduction

The lexico-optimal flow problem in a multiterminal network considered by N. Megiddo [6] may be described as follows. Let $N = (V, A, c; S^+, S^-)$ be a capacitated multiterminal network with vertex set V , arc set A , nonnegative capacity $c \in (\mathbf{R}_+ \cup \{0\})^A$ (where \mathbf{R}_+ denotes the set of positive reals), set of sources $S^+ (\neq \emptyset)$, and set of sinks $S^- (\neq \emptyset)$, where $S^+ \cup S^- \subseteq V$ and $S^+ \cap S^- = \emptyset$. A vector $\varphi \in \mathbf{R}^A$ is a flow in N if

$$0 \leq \varphi(a) \leq c(a) \quad (a \in A),$$

$$\begin{aligned} \sum_{a \in \delta^+ v} \varphi(a) - \sum_{a \in \delta^- v} \varphi(a) &= 0 \quad (v \in V - (S^+ \cup S^-)), \\ s^+(v) &\equiv \sum_{a \in \delta^+ v} \varphi(a) - \sum_{a \in \delta^- v} \varphi(a) \geq 0 \quad (v \in S^+), \\ s^-(v) &\equiv \sum_{a \in \delta^- v} \varphi(a) - \sum_{a \in \delta^+ v} \varphi(a) \geq 0 \quad (v \in S^-), \end{aligned}$$

where $\delta^+ v$ (resp. $\delta^- v$) is the set of all arcs having v as their initial (resp. terminal) vertex.

For a nonnegative vector $x \in (\mathbf{R}_+ \cup \{0\})^E$ and a positive vector $y \in \mathbf{R}_+^E$ (E : finite set) in general, we denote by $\omega(x, y)$ the $|E|$ -tuple of numbers $x(e)/y(e)$ ($e \in E$) arranged in order of increasing magnitude. The lexico-optimal flow problem with respect to given positive weight vectors $w^+ \in \mathbf{R}_+^{S^+}$ and $w^- \in \mathbf{R}_+^{S^-}$ is then to find a maximal flow φ such that both $\omega(s^+, w^+)$ and $\omega(s^-, w^-)$ are lexicographically largest possible. (The case of $w^+(v) = 1$ ($v \in S^+$) and $w^-(v) = 1$ ($v \in S^-$) is treated in [6].) This problem is generalized by S. Fujishige [4] to the lexico-optimal base problem for a polymatroid and an efficient algorithm is given in [4].

Put

$$f_\alpha(t) = \begin{cases} -\log t & (\alpha = -1); \\ \frac{4}{1-\alpha^2} (1 - t^{\frac{1+\alpha}{2}}) & (-1 < \alpha < 1); \\ t \log t & (\alpha = 1) \end{cases}$$

and for positive vectors $x, y \in \mathbf{R}_+^E$ we define

$$(1.1) \quad D_\alpha(x, y) = \sum_{e \in E} x(e) f_\alpha\left(\frac{y(e)}{x(e)}\right).$$

When $x(E) = y(E) = 1$, we may regard x and y as probability distributions on E , and then $D_\alpha(x, y)$ is called the α -divergence or the Chernoff distance of degree α ; D_{-1} agrees with the Kullback-Leibler divergence and D_0 with the Hellinger distance [1]. As a generalization of (1.1) we consider

$$(1.2) \quad D(x, y) = \sum_{e \in E} x(e) f\left(\frac{y(e)}{x(e)}\right)$$

with $f(t)$ being smooth and strictly convex for $t > 0$, and adopt D as a distance measure. When $x(E) = y(E) = 1$, $D(x, y)$ is called the f -divergence [2].

The problems treated here are polymatroidal generalizations of the following network-flow problems for N , where we further assume that a one-to-one correspondence is given between S^+ and S^- .

Problem N1: Find a maximal flow φ in N such that the supply vector s^+ and the demand vector s^- are as close as possible in the sense that $D(s^+, s^-)$ is smallest;

Problem N2: Find a maximal flow φ in N such that the supply vector s^+ and the demand vector s^- are as close as possible in the sense that $\omega(s^+, s^-)$ is largest.

The polymatroidal generalizations are as follows. Suppose we are given a pair of polymatroids (E, ρ_1) and (E, ρ_2) defined on the same ground set E with respective rank functions ρ_1 and ρ_2 ; the base polyhedra will be denoted by $B(\rho_1)$ and $B(\rho_2)$.

Problem M1: Minimize $D(x, y)$ subject to $x \in B(\rho_1), y \in B(\rho_2)$;

Problem M2: Maximize $\omega(x, y)$ subject to $x \in B(\rho_1), y \in B(\rho_2)$.

We call (x, y) a *nearest pair* with respect to D if it is a solution to **M1**, and a *lexico-optimal pair* if it is a solution to **M2**.

This note shows that the solution sets to the two problems coincide, and that they agree with the set of universal pairs (see §2).

2. Result

We shall largely follow the notation of [4]. For a finite set E , vectors $x, y \in \mathbf{R}^E$ and $S \subseteq E$ in general, $x(S)$ will mean $\sum_{e \in S} x(e)$ and $x \wedge y$ is a vector $z \in \mathbf{R}^E$ such that $z(e) = \min(x(e), y(e))$ for $e \in E$. For positive vectors $x, y \in \mathbf{R}_+^E$ let $c_1 < c_2 < \dots < c_p$ be the distinct numbers among $(x(e)/y(e) | e \in E)$ and put $S_k = S_k(x, y) = \{e \in E | x(e)/y(e) \leq c_k\}$ ($k = 1, \dots, p$); we set $c_0 = -\infty, c_{p+1} = +\infty, S_0(x, y) = \emptyset$.

For $i = 1, 2$, the set of independent vectors of polymatroid (E, ρ_i) is denoted by $P(\rho_i)$, whereas the set of bases, i.e., the base polyhedron, is by $B_i = B(\rho_i)$; $\text{dep}_i(x, e)$ will denote the dependence function of (E, ρ_i) . The polymatroid intersection problem for (E, ρ_i) ($i = 1, 2$) is to find a vector $x \in P(\rho_1) \cap P(\rho_2)$ that maximizes $x(E)$. A pair of bases $x \in B(\rho_1)$ and $y \in B(\rho_2)$ is called a *universal pair of bases* if $(1 - \theta)x \wedge (1 + \theta)y$ is a solution to the polymatroid intersection problem for $(E, (1 - \theta)\rho_1)$ and $(E, (1 + \theta)\rho_2)$ for all θ ($-1 < \theta < 1$) [9], [10], [11]. For each θ

$$L(\theta) \equiv \{S \subseteq E | \mu(S; \theta) \leq \mu(T; \theta), \forall T \subseteq E\}$$

constitutes a sublattice of 2^E , where

$$\mu(S; \theta) \equiv (1 - \theta)\rho_1(S) + (1 + \theta)\rho_2(E - S).$$

Furthermore,

$$L_{\text{all}} \equiv \bigcup_{-1 < \theta < 1} L(\theta)$$

is also a sublattice of 2^E ([5], [9], [10]).

The following theorem characterizes the solution to problems **M1** and **M2**. The equivalence (a) \iff (b) is well known; the equivalence (f) \iff (g) is established in [9], [10]; and the implication (f) \implies (d) is mentioned in [11] without proof.

Theorem. Assume $B(\rho_1) \subset \mathbf{R}_+^E$ and $B(\rho_2) \subset \mathbf{R}_+^E$, where \mathbf{R}_+ denotes the set of positive reals. For $x \in B(\rho_1)$ and $y \in B(\rho_2)$ the following seven conditions are equivalent.

(a)

$$\text{dep}_1(x, e) \subseteq S_k(x, y) \quad (\forall e \in S_k(x, y); k = 1, \dots, p),$$

$$\text{dep}_2(y, e) \subseteq E - S_k(x, y) \quad (\forall e \in E - S_k(x, y); k = 1, \dots, p).$$

(b)

$$\begin{aligned} x(S_k(x, y)) &= \rho_1(S_k(x, y)) \quad (k = 1, \dots, p), \\ y(E - S_k(x, y)) &= \rho_2(E - S_k(x, y)) \quad (k = 1, \dots, p). \end{aligned}$$

(c) x is the lexico-optimal base with respect to y (in the sense of [4]), and y is the lexico-optimal base with respect to x .

(d) (x, y) is a lexico-optimal pair.

(e) (x, y) is a nearest pair with respect to D .

(f) (x, y) is a universal pair.

(g)

$$\begin{aligned} x(S) &= \rho_1(S) \quad (\forall S \in L_{\text{all}}), \\ y(E - S) &= \rho_2(E - S) \quad (\forall S \in L_{\text{all}}). \end{aligned}$$

Proof: Let $\chi_e \in \mathbf{R}^E$ be such that $\chi_e(e) = 1$ and $\chi_e(e') = 0$ for $e' \neq e$.

(a) \iff (b): The equivalence of (a) and (b) is well known; see [3],[4].

(b) \iff (c): See [4].

(d) \Rightarrow (a): If $\exists e \in S_k, \exists e' \in \text{dep}_1(x, e) - S_k$, then $x' \equiv x + d(\chi_e - \chi_{e'})$ belongs to B_1 for some $d > 0$ and $\omega(x', y) > \omega(x, y)$. This establishes the first assertion of (a). Similarly for the second.

(b) \Rightarrow (d): Put $T_0 = \emptyset, T_k = S_k(x, y) - S_{k-1}(x, y) \quad (k = 1, \dots, p)$ and let (\hat{x}, \hat{y}) be a lexico-optimal pair. We shall show by induction on $k = 0, \dots, p$ that

$$x(e)/y(e) = \hat{x}(e)/\hat{y}(e) \quad (e \in T_k).$$

This is trivially true for $k = 0$.

Suppose that

$$x(e)/y(e) = \hat{x}(e)/\hat{y}(e) \quad (e \in S_{k-1}).$$

Since $\omega(x, y) \leq \omega(\hat{x}, \hat{y})$, we must have

$$(2.1) \quad x(e)/y(e) = c_k \leq \hat{x}(e)/\hat{y}(e) \quad (e \in T_k).$$

From this follows that

$$(2.2) \quad c_k \hat{y}(T_k) \leq \hat{x}(T_k).$$

On the other hand, the assumption on (x, y) and the optimality of (\hat{x}, \hat{y}) together yield

$$x(T_k) = \hat{x}(T_k) (= \rho_1(S_k) - \rho_1(S_{k-1})),$$

$$y(T_k) = \hat{y}(T_k) (= \rho_2(E - S_{k-1}) - \rho_2(E - S_k)),$$

where the established implication (d) \Rightarrow (b) is used. Substituting these relations into (2.2) and noting the fact $c_k = x(T_k)/y(T_k)$ we obtain

$$x(T_k) = c_k y(T_k) = c_k \hat{y}(T_k) \leq \hat{x}(T_k) = x(T_k).$$

This shows that (2.2) and (2.1) hold with equalities.

(a) \iff (e): Put

$$(2.3) \quad g(u, v) = uf(v/u),$$

which is convex in $(u, v) \in \mathbf{R}_+^2$ on account of the convexity of f and has the derivatives $g_u \equiv \partial g/\partial u, g_v \equiv \partial g/\partial v$ given as follows:

$$(2.4) \quad g_u = f\left(\frac{v}{u}\right) - \frac{v}{u}f'\left(\frac{v}{u}\right), \quad g_v = f'\left(\frac{v}{u}\right).$$

From these expressions and the strict convexity of f , we see that

$$(2.5) \quad g_u(e') \leq g_u(e) \iff g_v(e') \geq g_v(e) \iff \frac{x(e')}{y(e')} \leq \frac{x(e)}{y(e)},$$

where $g_u(e) \equiv g_u(x(e), y(e))$ and $g_v(e) \equiv g_v(x(e), y(e))$. Namely, (a) is equivalent to (2.6) below:

$$(2.6a) \quad g_u(e') \leq g_u(e) \quad (e' \in \text{dep}_1(x, e)),$$

$$(2.6b) \quad g_v(e') \geq g_v(e) \quad (e \in \text{dep}_2(y, e')).$$

Suppose (x, y) minimizes $D(x, y)$. Then we must have (2.6), since

$$D(x, y) = \sum_{e \in E} g(x(e), y(e)).$$

Conversely suppose (a) (and hence (2.6)) holds. Let (\hat{x}, \hat{y}) ($\hat{x} \in B_1, \hat{y} \in B_2$) give the minimum of $D(x, y)$. Since $x, \hat{x} \in B_1$, there exist $d_1(e'', e') \geq 0$ ($e' \in \text{dep}_1(x, e'')$) (see [3]) such that

$$\hat{x} = x + \sum \{d_1(e'', e')(\chi_{e''} - \chi_{e'}) \mid e' \in \text{dep}_1(x, e'')\}$$

(where the summation is taken over all pairs (e'', e') with the indicated relation), i.e.,

$$\hat{x}(e) = x(e) + \sum \{d_1(e, e') \mid e' \in \text{dep}_1(x, e)\} - \sum \{d_1(e'', e) \mid e \in \text{dep}_1(x, e'')\}.$$

Similarly there exist $d_2(e'', e') \geq 0$ ($e' \in \text{dep}_2(y, e'')$) such that

$$\hat{y}(e) = y(e) + \sum \{d_2(e, e') \mid e' \in \text{dep}_2(y, e)\} - \sum \{d_2(e'', e) \mid e \in \text{dep}_2(y, e'')\}.$$

These relations, combined with the convexity of $g(u, v)$, yield

$$\begin{aligned} \hat{g}(e) \geq g(e) + g_u(e) & \left[\sum_{e': e' \in \text{dep}_1(x, e)} d_1(e, e') - \sum_{e'': e \in \text{dep}_1(x, e'')} d_1(e'', e) \right] \\ & + g_v(e) \left[\sum_{e': e' \in \text{dep}_2(y, e)} d_2(e, e') - \sum_{e'': e \in \text{dep}_2(y, e'')} d_2(e'', e) \right], \end{aligned}$$

where $\hat{g}(e) \equiv g(\hat{x}(e), \hat{y}(e))$ and $g(e) \equiv g(x(e), y(e))$. Taking the sum of the above expressions over $e \in E$, we obtain

$$D(\hat{x}, \hat{y}) - D(x, y) \geq \sum_{e' \in \text{dep}_1(x, e)} [g_u(e) - g_u(e')]d_1(e, e') + \sum_{e \in \text{dep}_2(y, e')} [g_v(e') - g_v(e)]d_2(e', e),$$

which is nonnegative by (2.6). (The summations are taken over all pairs (e, e') with the indicated relations.)

(b) \iff (f): By choosing $\theta = (c_k - 1)/(c_k + 1)$ we can easily show that (f) implies (b). Conversely suppose (b) holds. For any θ there exists $k (0 \leq k \leq p)$ such that $c_k \leq (1 + \theta)/(1 - \theta) < c_{k+1}$. The relation

$$\begin{aligned} [(1 - \theta)x \wedge (1 + \theta)y](E) &= (1 - \theta)x(S_k) + (1 + \theta)y(E - S_k) \\ &= (1 - \theta)\rho_1(S_k) + (1 + \theta)\rho_2(E - S_k) \end{aligned}$$

implies (f) when combined with the well-known minimax relation for polymatroid intersection problem. See [9], [10] for further details.

(f) \iff (g): See [9], [10]. \square

3. Remarks

Remark 1: The theorem implies that the problems **M1** and **M2** can be solved efficiently by utilizing the algorithm of [9], [10] designed for the universal pair.

Remark 2: The theorem does not apply to integer versions of the problems **M1** and **M2** for integral polymatroids.

Remark 3: The theorem reveals that the nearest pair with respect to D_α for a particular value of α is universally nearest for all α .

Remark 4: We have observed that g of the form (2.3) with strictly convex f enjoys the crucial property (2.5). The converse is also true as follows, which would justify our assumption (1.2) on the form of D . Suppose $g(u, v)$ is a smooth function having the property (2.5). Then there are smooth functions, say, $\varphi(t)$ and $\psi(t)$ defined for $t > 0$ such that

$$\begin{aligned} g_u(u, v) &= \varphi(v/u), & \varphi'(t) &\leq 0 (t > 0), \\ g_v(u, v) &= \psi(v/u), & \psi'(t) &\geq 0 (t > 0). \end{aligned}$$

Putting $G(r, \theta) = g(r \cos \theta, r \sin \theta)$ we see that both $\partial G/\partial r$ and $(1/r)\partial G/\partial \theta$ are independent of r . This implies that

$$G(r, \theta) = rH(\theta) + c$$

with a function $H(\theta)$ and a constant c , or equivalently,

$$g(u, v) = \sqrt{u^2 + v^2} h(v/u) + c$$

with a function $h(t)$ and a constant c . We see that $g(u, v)$ can be written in the form (2.3) by putting $f(t) = \sqrt{1+t^2} h(t)$ and assuming $c = 0$ (without loss of generality since $x(E)$ is constant for $x \in B_1$). Then the convexity of $f(t)$ follows from (2.4) and (2.5).

Remark 5: In §1, we have mentioned a network $N = (V, A, c; S^+, S^-)$ with a one-to-one correspondence between S^+ and S^- . For N we may think of the following two optimization problems:

(P1) Maximize the total amount of flow $s^+(S^+) (= s^-(S^-))$;

(P2) Minimize (lexicographically) $\omega(s^+, s^-)$.

If we consider (P2) among the solutions to (P1), we are led to the problem treated in this paper. On the other hand, if we consider (P1) among the solutions to (P2), we are led to the problem of maximizing the total amount of a flow subject to $s^+ = s^-$; the maximum value is known to be equal to the maximum amount of a flow in the product (i.e., the cascade connection) of many copies of N (see [7], Theorem 4.3 of [8]).

Acknowledgement

The author thanks Professor Satoru Fujishige for giving a stimulating comment on the earlier manuscript of this paper and for informing him of the relevant reference [11]. Some of the comments by the anonymous referees are also helpful in improving this paper.

References

- [1] Amari, S.: *Differential-Geometrical Methods in Statistics*. Lecture Notes in Statistics, **28**, Springer-Verlag, 1985.
- [2] Csiszár, I.: Information-type Measures of Difference of Probability Distributions and Indirect Observations. *Studia Scientiarum Mathematicarum Hungarica*, **2** (1967), 299-318.
- [3] Fujishige, S. : Algorithms for Solving the Independent-Flow Problems. *Journal of the Operations Research Society of Japan*, **21** (1978), 189-203.
- [4] Fujishige, S. : Lexicographically Optimal Base of a Polymatroid with Respect to a Weight Vector. *Mathematics of Operations Research*, **5**(1980), 186-196.
- [5] Iri, M.: Structural Theory for the Combinatorial Systems Characterized by Submodular Functions. In *Progress in Combinatorial Optimization* (ed. W. R. Pulleyblank), Academic Press, 197-219, 1984.
- [6] Megiddo, N.: Optimal Flows in Networks with Multiple Sources and Sinks. *Mathematical Programming*, **7** (1974), 97-107.
- [7] Murota, K.: Combinatorial Dynamical System Theory. Discussion Paper Series 317, Institute of Socio-Economic Planning, University of Tsukuba, 1986. Revised version: Research Memorandum RMI 86-02, Department of Mathematical Engineering and

- Information Physics, University of Tokyo, 1986. Also *Discrete Applied Mathematics*, to appear.
- [8] Murota, K.: Eigensets and Power Products of a Bimatroid. *Advances in Mathematics*, to appear.
- [9] Nakamura, M.: *Mathematical Analysis of Discrete Systems and Its Applications* (in Japanese). Doctor's dissertation, Department of Mathematical Engineering and Information Physics, University of Tokyo, 1983.
- [10] Nakamura, M., and Iri, M.: A Structural Theory for Submodular Functions, Polymatroids and Polymatroid Intersections. Research Memorandum RMI 81-06, Department of Mathematical Engineering and Information Physics, University of Tokyo, 1981.
- [11] Tomizawa, N.: Theory of Hyperspaces (IV)- Principal Partitions of Hypermatroids (in Japanese). Technical Report of Institute of Electronics and Communication Engineers of Japan, CAS 80-85, 17-24 (1980).

Kazuo Murota: Department of Mathematical
Engineering and Information Physics, Faculty of Engineering, University of Tokyo,
Bunkyo-ku, Tokyo 113, Japan