Finding a subset of nonnegative vectors with a coordinatewise large sum[☆]

Ilya I. Bogdanov[a,b,*], Grigory R. Chelnokov[b]

[a] *Moscow Institute of Physics and Technology (State University), Institutsky per., 9, Dolgoprudny, Moscow reg., Russia 141700*
[b] *Laboratory of Discrete and Computational Geometry, Yaroslavl' State University, Sovetskaya st. 14, Yaroslavl', Russia 150000*

Abstract

Given a rational $a = p/q$ and N nonnegative d-dimensional real vectors $\mathbf{u}_1$, ..., $\mathbf{u}_N$, we show that it is always possible to choose $(d-1)+\lceil(pN-d+1)/q\rceil$ of them such that their sum is (componentwise) at least $(p/q)(\mathbf{u}_1+\cdots+\mathbf{u}_N)$. For fixed d and a, this bound is sharp if N is large enough. The method of the proof uses Carathéodory's theorem from linear programming.

Keywords: subsum optimization, linear programming

1. Introduction

We deal with the d-dimensional real vector space $\mathbb{R}^d$; the vectors of the standard basis are denoted by $\mathbf{e}_1, \ldots, \mathbf{e}_d$. Introduce a coordinatewise partial order $\succeq$ on $\mathbb{R}^d$; that is, for the vectors $\mathbf{u} = [u^1, \ldots, u^d]$ and $\mathbf{v} = [v^1, \ldots, v^d]$ we write $\mathbf{u} \succeq \mathbf{v}$ if $u^j \geq v^j$ for $1 \leq j \leq d$.

Let $\mathbf{u}_1, \ldots, \mathbf{u}_N \in \mathbb{R}^d$ be N nonnegative vectors (that is, $\mathbf{u}_i \succeq \mathbf{0}$ for $1 \leq i \leq N$), and let $a \in [0,1]$ be some real number. We say that a set of indices $I \subseteq \{1, \ldots, N\}$ is *a-rich* if

$$\sum_{i \in I} \mathbf{u}_i \succeq a \sum_{i=1}^{N} \mathbf{u}_i.$$

[☆]Supported by the Russian government project 11.G34.31.0053.

[*]Corresponding author

Email addresses: ilya.i.bogdanov@gmail.com (Ilya I. Bogdanov), grishabenruven@yandex.ru (Grigory R. Chelnokov)

Let $f_{N,d}(a)$ be the minimal number f such that for every N nonnegative vectors $\mathbf{u}_1, \dots, \mathbf{u}_N \in \mathbb{R}^d$ there exists an a-rich set I with $|I| \le f$. Further we consider only rational a and write $a = p/q$ with $q > 0$ and $\gcd(p, q) = 1$.

To find an upper bound for $f_{N,d}(a)$, one may use a theorem of Stromquist and Woodall [1] claiming that, given n non-atomic probability measures on S^1, there exists a union of $n-1$ arcs that has measure a in each measure. It can be performed as follows. Let $w^j = \sum_{i=1}^N u_i^j$; we may assume that $w_j > 0$ for $1 \le j \le d$. Consider a segment $T = [0, N]$, identify its endpoints to obtain a circle of length N, and split it into unit segments. For $1 \le i \le N$ and $1 \le j \le d$, define a measure μ^j on segment $[i-1, i]$ as $\mu^j = u_i^j \mu / w^j$, where μ is the usual Lebesgue measure; set also $\mu^{d+1} = \mu/N$. By the theorem mentioned above, there exists a union of d arcs $J \subseteq T$ such that $\mu^j(\mathcal{F}) = a$ for $1 \le j \le d+1$. Now, one may define

$$I = \{i : J \cap [i-1, i] \neq \varnothing\}.$$

This set is a-rich since

$$\sum_{i \in I} u_i^j = w^j \mu^j \left(\bigcup\nolimits_{i \in I} [i-1, i] \right) \ge w^j \mu^j(J) = a \sum_{i=1}^{n} u_i^j.$$

Moreover,

$$|I| \le \mu(J) + 2d = N\mu^{d+1}(J) + 2d = aN + 2d.$$

Thus, $f_{N,d}(a) \le aN + 2d$.

In an analogous way, one may apply a well-known Alon's theorem on splitting of necklaces [2] obtaining a bound

$$f_{N,d}(p/q) \le \frac{p}{q} \cdot N + \frac{p(q-p)}{q} \cdot d.$$

The bounds shown above are asymptotically tight. Nevertheless, they provide exact values of $f_{N,d}(p/q)$ only for some border cases. The aim of this paper is to find an exact value of $f_{N,d}(a)$ for every rational $a = p/q$, positive integer d and sufficiently large N. We use only the methods of linear programming.

The main result is the following theorem.

Theorem 1. *For any positive integer numbers N, d and rational number $a = p/q \in [0, 1]$, we have*

$$f_{N,d}(p/q) \le (d-1) + \left\lceil \frac{pN - d + 1}{q} \right\rceil.$$

Moreover, if $q > p \geq 1$ and $N \geq (q-1)(d-1)$, then we have

$$f_{N,d}(p/q) = (d-1) + \left\lceil \frac{pN - d + 1}{q} \right\rceil.$$

Throughout the rest of the paper, we use the notation $s = d - 1$.

The next section contains the proof of the upper bound. Here we present an example showing that this bound is sharp if N is large enough.

Example 1. Choose an integer $r \in [1, q-1]$ such that $pr \equiv 1 \pmod{q}$. Let $m = \lceil pr/q \rceil$ (hence $qm - pr = q - 1$).

Let us set $\mathbf{u}_{ir-k} = \mathbf{e}_i$ for $1 \leq i \leq s$ and $0 \leq k \leq r-1$, and set $\mathbf{u}_i = \mathbf{e}_d$ for $i > rs$ (notice that $N \geq (q-1)s \geq rs$). Denote $\mathbf{w} = \sum_{i=1}^N \mathbf{u}_i = [r, r, \ldots, r, N - rs]$. Now, if $\sum_{i \in I} \mathbf{u}_i \succeq \frac{p}{q}\mathbf{w}$, then

$$\sum_{i \in I} \mathbf{u}_i \succeq \left[m, m, \ldots, m, \left\lceil \frac{p}{q}(N - rs) \right\rceil \right],$$

because all the coordinates of $\mathbf{u}_i$ are integer. Thus, since the sum of coordinates of each vector is 1, we should have

$$|I| \geq ms + \left\lceil \frac{p}{q}(N - rs) \right\rceil = s + \left\lceil \frac{pN + s(qm - pr - q)}{q} \right\rceil = s + \left\lceil \frac{pN - s}{q} \right\rceil,$$

as desired. $\square$

2. Proof of the upper bound

Consider N vectors $\mathbf{u}_1, \ldots, \mathbf{u}_N \in \mathbb{R}^d$ with nonnegative coordinates. Denote

$$\mathbf{w} = \sum_{i=1}^N \mathbf{u}_i, \qquad f = s + \left\lceil \frac{pN - s}{q} \right\rceil.$$

We need to prove that there exists a set of indices I such that

$$|I| \leq f \qquad \text{and} \qquad \sum_{i \in I} \mathbf{u}_i \succeq \frac{p}{q}\mathbf{w}.$$

We use induction on $p + q$. In the base cases $p + q \leq 2$ we have $a = 0$ or $a = 1$, and the statement is trivial.

Now, assume that $p+q \geq 3$, and assume that the statement of the theorem holds for all pairs (p', q') with $p' + q' < p + q$. Introduce the following set:

$$X_{p/q} = \left\{ \mathbf{x} \in \mathbb{R}^N : \quad 0 \leq x_i \leq 1, \quad \sum_{i=1}^{N} x_i \mathbf{u}_i = \frac{p}{q}\mathbf{w} \right\}.$$

This set is closed and bounded; next, it is nonempty since $[p/q, \ldots, p/q] \in X_{p/q}$. Moreover, this set is defined by $s+1$ linear equalities and some linear inequalities. By Carathéodory's theorem, there exists a vector $\mathbf{x} = [x_1, \ldots, x_N] \in X_{p/q}$ such that $N-s-1$ of these inequalities come to equalities; that is, $N-s-1$ coordinates of $\mathbf{x}$ are integer. Hence, either (i) at least $N-f$ coordinates are zeros, or (ii) at least $f-s$ coordinates are ones.

In case (i), denote $I = \{i : x_i > 0\}$. We have $|I| \leq N - (N-f) = f$. On the other hand, we obtain

$$\sum_{i \in I} \mathbf{u}_i \succeq \sum_{i \in I} x_i \mathbf{u}_i = \sum_{i=1}^{N} x_i \mathbf{u}_i = \frac{p}{q}\mathbf{w}, \tag{1}$$

as desired.

In case (ii), define $J = \{i : x_i < 1\}$, and let $N' = |J|$. We have

$$\sum_{i \in J} \mathbf{u}_i = \mathbf{w} - \sum_{i \notin J} \mathbf{u}_i \succeq \mathbf{w} - \sum_{i=1}^{N} x_i \mathbf{u}_i = \frac{q-p}{q}\mathbf{w}. \tag{2}$$

Notice that in (1) and (2) we have used the condition that all the vectors $\mathbf{u}_i$ are nonnegative.

Next, note that

$$N' = |J| \leq N - f + s = N - \left\lceil \frac{pN - s}{q} \right\rceil \leq N - \frac{pN - s}{q} = \frac{(q-p)N + s}{q}. \tag{3}$$

Renumbering the vectors we may assume that $J = \{1, 2, \ldots, N'\}$. Again, we distinguish two subcases: (ii') $q \geq 2p$ and (ii'') $q < 2p$.

In case (ii'), we apply the induction hypothesis to the vectors $\mathbf{u}_1, \ldots, \mathbf{u}_{N'}$ and the number $a' = p/(q-p) \in [0, 1]$. We obtain the subset $I \subseteq \{1, \ldots, N'\}$ such that

$$|I| \leq s + \left\lceil \frac{pN' - s}{q-p} \right\rceil$$

and

$$\sum_{i\in I}\mathbf{u}_i \succeq \frac{p}{q-p}\sum_{i\in J}\mathbf{u}_i.$$

By (2), the last inequality yields

$$\sum_{i\in I}\mathbf{u}_i \succeq \frac{p}{q-p}\cdot\frac{q-p}{q}\mathbf{w} = \frac{p}{q}\mathbf{w}. \tag{4}$$

Next, by (3) we have

$$pN' - s \le \frac{p}{q}\big((q-p)N + s\big) - s = \frac{q-p}{q}(pN-s),$$

thus obtaining

$$|I| \le s + \left\lceil \frac{1}{q-p}\cdot\frac{q-p}{q}(pN-s)\right\rceil = f. \tag{5}$$

The relations (4) and (5) show that I is a desired set of indices.

In case (ii''), we apply the induction hypothesis to $N-N'$ vectors $\mathbf{u}_{N'+1}$, $\ldots$, $\mathbf{u}_N$ and the number $a' = (2p-q)/p \in (0,1)$. We obtain the subset $I' \subseteq \{N'+1,\ldots,N\}$ such that

$$|I'| \le s + \left\lceil \frac{(2p-q)(N-N')-s}{p}\right\rceil$$

and

$$\sum_{i\in I'}\mathbf{u}_i \succeq \frac{2p-q}{p}\sum_{i=N'+1}^{N}\mathbf{u}_i.$$

Now we claim that the subset $I = I'\cup J$ satisfies the desired properties. Recall that by (2) we have

$$\sum_{i=1}^{N'}\mathbf{u}_i = \frac{q-p}{q}\mathbf{w} + \mathbf{w}'$$

for some vector $\mathbf{w}' \succeq \mathbf{0}$. Hence

$$\sum_{i=N'+1}^{N}\mathbf{u}_i = \frac{p}{q}\mathbf{w} - \mathbf{w}',$$

and we obtain

$$\sum_{i\in I}\mathbf{u}_i=\sum_{i\in I'}\mathbf{u}_i+\sum_{i\in J}\mathbf{u}_i\succeq\frac{2p-q}{p}\left(\frac{p}{q}\mathbf{w}-\mathbf{w}'\right)+\left(\frac{q-p}{q}\mathbf{w}+\mathbf{w}'\right)$$
$$=\frac{p}{q}\mathbf{w}+\frac{q-p}{p}\mathbf{w}'\succeq\frac{p}{q}\mathbf{w}.$$

We are left to show that $|I|\le f$.

Recall that

$$|I|=|J|+|I'|\le N'+s+\left\lceil\frac{(2p-q)(N-N')-s}{p}\right\rceil$$
$$=s+\left\lceil\frac{(q-p)N'+(2p-q)N-s}{p}\right\rceil.$$

So, it suffices to prove that

$$\frac{(q-p)N'+(2p-q)N-s}{p}\le\frac{pN-s}{q},\qquad\text{or}\qquad qN'\le(q-p)N+s,$$

which is equivalent to (3). Thus I is a desired set. $\square$

Acknowledgements

The authors are grateful to Prof. V.L. Dolnikov and Prof. R.N. Karasev for useful discussions. The authors thank the referees for their valuable remarks.

References

[1] W. Stromquist and D. Woodall, Sets on which several measures agree, J. Math. Anal. Appl., 108, 1985, 241–248.

[2] N. Alon, Splitting of necklaces, Adv. in Math., 63(3), 1987, 247–253.