Les Cahiers du GERAD ISSN: 0711–2440

A lower bound on the sum of the largest components of a nonnegative vector

C. Audet

G–2015–54

May 2015

La publication de ces rapports de recherche est rendue possible grâce au soutien de HEC Montréal, Polytechnique Montréal, Université McGill, Université du Québec à Montréal, ainsi que du Fonds de recherche du Québec – Nature et technologies.

Dépôt légal – Bibliothèque et Archives nationales du Québec, 2015.

The publication of these research reports is made possible thanks to the support of HEC Montréal, Polytechnique Montréal, McGill University, Université du Québec à Montréal, as well as the Fonds de recherche du Québec – Nature et technologies.

Legal deposit – Bibliothèque et Archives nationales du Québec, 2015.

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

A lower bound on the sum of the largest components of a nonnegative vector

Charles Audet

GERAD & Département de mathématiques et de génie industriel, Polytechnique Montréal, Montréal (Québec) Canada, H3C 3A7

charles.audet@gerad.ca

May 2015

Les Cahiers du GERAD

G–2015–54

Abstract: We study the function returning the sum of the k components of largest magnitude of a vector. We show that if a nonnegative vector x is such that its Euclidean norm is greater than or equal to one, and if the integer k is an upper bound on the Manhattan norm of x, then there exists k components of x whose sum is greater than or equal to one.

Key Words: Sum of largest components of a vector, order statistics.

Acknowledgments: Work supported by NSERC grant 2015-05311.

1 Introductory note

We consider the function that returns the sum of the largest k components of a vector in $\mathbb{R}^n$, where k is an integer between 1 and n. Formally, define the partial sum function $s_k : \mathbb{R}^n \to \mathbb{R}$ as

$$s_k(x) := \sum_{i=1}^{k} x_{[i]}$$

where $x_{[i]}$ denotes the i-th largest component of the vector $x \in \mathbb{R}^n$. When sorted in increasing order, the values $x_{[n]} \leq x_{[n-1]} \leq \ldots \leq x_{[1]}$ are know as order statistics [3]. The textbooks [1, 2] show that this partial sum function s_k is convex by using elementary linear programming duality theory.

In the present note, we show that any $x \in \mathbb{R}^n_+$ whose Euclidean and Manhattan norms satisfy $\|x\|_2 \geq 1$ and $\|x\|_1 \leq k$ for some integer $k \leq n$, is such that the sum of the k largest components of x is at least one, i.e., $s_k(x) \geq 1$. The inequality is trivial when $k = 1$ because the only admissible vectors are the coordinate vectors. The inequality is also trivial when $k = n$ because $s_n(x) = \|x\|_1 \geq \|x\|_2$ for any nonnegative vector x.

2 Proof of the main result

The following Lemma is used twice in the demonstration of the main result.

Lemma 2.1 *Let $m \in \mathbb{N}$, $M = \{1, 2, \ldots, m\}$, and α, β, γ be nonnegative real numbers satisfying $\alpha \leq \beta$. If a and b belong to $\Omega = \{c \in \mathbb{R}^n : \|c\|_1 = \gamma,\ \alpha \leq c_i \leq \beta\ \forall i \in M\}$, and if all the components of a, except for possibly one, are at their bound α or β then $b^T b \leq a^T a$.*

Proof. Let a and b in Ω be such that the components of a, except for possibly one, are either α or β. Choose $c \in \Omega$ so that $c^T c \geq d^T d$ for all $d \in \Omega$. In particular, we have $c^T c \geq b^T b$.

If there exists two indices $i \neq j$ of M such that $\alpha < c_i \leq c_j < \beta$, then $\delta := \min\{c_i - \alpha, c_j + \delta\} > 0$ satisfies $\alpha \leq c_i - \delta < c_j + \delta \leq \beta$ and

$$(c_i - \delta)^2 + (c_j + \delta)^2 = c_i^2 + c_j^2 + 2\delta(c_j - c_i) + 2\delta^2 > c_i^2 + c_j^2.$$

This situation would contradict the fact that $c^T c$ is maximal. Therefore, at most one component of c differs from α and β.

Furthermore, since $\|a\|_1 = \|c\|_1$ and $\alpha \leq a_i \leq \beta$ et $\alpha \leq c_i \leq \beta$ for all $i \in M$, it follows that a and c have the same entries, with possibly a different order. We conclude that $a^T a = c^T c \geq b^T b$. □

We can now state and prove the principal result.

Theorem 2.2 *If $k \leq n$ is an integer such that $k \geq \|x\|_1$ for some $x \in \mathbb{R}^n_+$ with $\|x\|_2 \geq 1$, then $s_k(x) \geq 1$.*

Proof. Consider the pair $x \in \mathbb{R}^n_+$ and $k \in \mathbb{N}$ with $\|x\|_2 \geq 1$ and $k \geq \|x\|_1$. Without any loss in generality, and by reordering the components if necessary, assume that $x_1 \geq x_2 \geq \ldots \geq x_n \geq 0$ and that $1 \leq k < n$. Furthermore, the case where $x_k = 0$ is trivial.

Set $v = x_k > 0$, and by an Euclidian division, find $p \in \mathbb{N}$ and $0 \leq q < v$ such that

$$pv + q = \sum_{i=k+1}^{n} x_i = \|x\|_1 - s_k(x).$$

Construct the vector $y \in \mathbb{R}^n_+$ as follows :

$$y_i = \begin{cases} s_k(x) - (k-1)v & \text{if } i = 1 \\ v & \text{if } 2 \leq i \leq p + k \\ q & \text{if } i = p + k + 1 \\ 0 & \text{if } p + k + 2 \leq i \leq n. \end{cases}$$

With this construction, we have $s_k(x) = s_k(y)$, and in addition, the one-norm of y satisfies

$$\begin{array}{rcl} \|y\|_1 \;=\; y_1 + (p+k-1)v + q & = & y_1 + (k-1)v + (pv+q) \\ & = & \sum_{i=1}^{k} x_i + \sum_{i=k+1}^{n} x_i \\ & = & \|x\|_1 \;\leq\; k. \end{array}$$

Multiplying this last inequality by the scalar $v > 0$, and by reordering the terms we get

$$-(pv+q)v \quad \geq \quad (y_1 - k + (k-1)v)v. \tag{1}$$

Let us now analyze the Euclidean norm of y. On the one hand, both vectors $(y_1, y_2, \ldots, y_k) = (y_1, v, v, \ldots, v)$ and $(x_1, x_2, \ldots, x_k)$ belong to the set $\Omega \subset \mathbb{R}^k$ of Lemma 2.1 with $\alpha = v \leq \beta = y_1$ and $\gamma = \sum_{i=1}^{k} x_i$. In addition, all components of the first vector are at their bounds and consequently

$$\sum_{i=1}^{k} x_i^2 \quad \leq \quad \sum_{i=1}^{k} y_i^2. \tag{2}$$

On the other hand, both vectors $(y_{k+1}, y_{k+2}, \ldots, y_n) = (v, v, \ldots, q, 0, 0, \ldots, 0)$ and $(x_{k+1}, x_{k+2}, \ldots, x_n)$ belong to $\Omega \subset \mathbb{R}^{n-k}$ of Lemma 2.1 with $\alpha = 0 < \beta = v$ et $\gamma = pv + q$. Once again, all components of the first vector, except possibly the one whose value is q, are at one of their bounds. It follows that

$$\sum_{j=k+1}^{n} x_j^2 \quad \leq \quad \sum_{j=k+1}^{n} y_j^2. \tag{3}$$

Combining Equations (2) and (3) with the assumption on the Euclidean norm of x leads to

$$\|y\|_2^2 \quad = \quad y_1^2 + (p+k-1)v^2 + q^2 \;\geq\; \|x\|_2^2 \;\geq\; 1, \tag{4}$$

and therefore

$$\begin{array}{rcll} y_1^2 + (k-1)y_1 v & \geq & y_1^2 + (k-1)v^2 & \text{since } y_1 \geq v \\ & \geq & 1 - pv^2 - q^2 & \text{by Equation (4)} \\ & \geq & 1 - (pv+q)v & \text{since } v \geq q \\ & \geq & 1 + (y_1 - k + (k-1)v)v & \text{by Equation (1).} \end{array} \tag{5}$$

Finally, consider the quantity

$$\begin{array}{rcll} (y_1 + 1 - v)(y_1 + (k-1)v - 1) & = & y_1^2 + (k-1)y_1 v - y_1 + (1-v)(y_1 + (k-1)v - 1) & \\ & \geq & 1 + (y_1 - k + (k-1)v)v & \\ & & \quad -y_1 + (1-v)(y_1 + (k-1)v - 1) & \text{by (5)} \\ & = & 1 + y_1 v - kv + (k-1)v^2 & \\ & & \quad -y_1 + y_1 + (k-1)v - 1 - y_1 v - (k-1)v^2 + v & \\ & = & 0. & \end{array}$$

However, $y_1 \geq v$ implies that both $(y_1 + 1 - v)$ and $(y_1 + (k-1)v - 1)$ are nonnegative, and

$$s_k(x) \;=\; s_k(y) \;=\; y_1 + (k-1)v \;\geq\; 1.$$

□

The Theorem may be trivially stated in $\mathbb{R}^n$ rather than $\mathbb{R}^n_+$ as follows. If $k \leq n$ is an integer such that $k \geq \|x\|_1$ for some $x \in \mathbb{R}^n$ with $\|x\|_2 \geq 1$, then $s_k(|x|) \geq 1$, where $|x|$ denotes the vector in $\mathbb{R}^n$ whose components are the absolute values of those of x.

As a final coment, observe that the Theorem cannot be made tighter since for any positive integer $k \leq \sqrt{n}$, the vector $x \in \mathbb{R}^n_+$ defined as

$$x_i = \begin{cases} \dfrac{1}{k} & \text{if } \ i \in \{1, 2, \ldots, k^2\} \\ 0 & \text{if } \ i \in \{k^2+1, k^2+2, \ldots, n\} \end{cases}$$

satisfies $\|x\|_2 = 1$, $\|x\|_1 = k$ and $s_k(x) = 1$.

References

[1] S.P. Boyd and L. Vandenberghe. Convex Optimization. Berichte über verteilte messysteme. Cambridge University Press, 2004.

[2] G.C. Calafiore and L. El Ghaoui. Optimization Models. Control systems and optimization series. Cambridge University Press, October 2014.

[3] H.A. David and H.N. Nagaraja. Order Statistics. Wiley Series in Probability and Statistics. Wiley, 2003.